*Corresponding Author Email: zijizhao@student.unimelb.edu.au

Characteristic atmospheric states during mid-summer droughts over Central America and Mexico

Zijie Zhao[1,2*], Neil J. Holbrook[3,4], Eric C. J. Oliver[4,5], Daniel Ballestero[6], J. Mauro Vargas-Hernandez[6]

1. School of Earth Sciences, The University of Melbourne, Melbourne, Victoria, Australia
2. Australian Research Council Centre of Excellence for Climate Extremes, The University of Melbourne, Melbourne, Victoria, Australia
3. Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia
4. Australian Research Council Centre of Excellence for Climate Extremes, University of Tasmania, Hobart, Tasmania, Australia
5. Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada
6. Universidad Nacional, Costa Rica, Departamento de Física.

Abstract

Annual precipitation over Central America and large areas of Mexico is typically characterised by its bimodal distribution, with a precipitation minimum in July to August that occurs between two separate maxima from May to July and August to October. Several theories have been proposed to explain this phenomenon, which is often termed the mid-summer drought (MSD), but most fail to address the different characteristics associated with individual MSD events. Here, a regression–based approach is used to detect and quantify the annual and climatological MSD signature over Central America and Mexico. This approach has been evaluated and shown to be robust for various datasets with different spatial resolutions. It was found that in the southeast of the Mexico/Central America region, MSDs start earlier and end later than elsewhere, and are thus longer in duration. However, the coast of the Gulf of Mexico, Cuba, and large areas of Central America, exhibit climatologically stronger MSDs. Changes in precipitation, brought about by the interaction between reversals

of the onshore/offshore winds and orographic forcing associated with the steep mountainous terrain, have also been shown to be significant factors in the timing of MSD occurrences, offering support for a combined theory of large-scale dynamics and regional forcing. Using self-organising maps (SOMs) as an analysis tool, it was found that MSD events over the domain display strong spatial variability. The MSDs over the domain also generate distinct signatures and may be forced by particular mechanisms. We found that El Niño-Southern Oscillation (ENSO) could be a potential classifier for the SOM identified atmospheric states, based on the correspondence of MSD occurrences with ENSO phases.

1 Introduction

Precipitation plays a significant role in a wide range of economic and other human activities throughout Central America and Mexico, including agriculture, power generation, and public health (Parmenter et al., 1999; Chowell and Sanchez, 2006; Brunkard et al., 2008; Fuchs and Wolff, 2011; Rogé et al., 2014). In particular, rainfall variations can substantially affect agriculture, including the production of corn and fruits that constitute an important part of the Gross National Product for many countries in this region. This also highlights the value of being able to accurately describe and quantify the spatio-temporal distribution of rainfall variability across the region as well as identifying the climatic factors that influence these changes, including near-surface temperatures and wind patterns, which may be potentially predictable.

The annual cycle of precipitation in most regions of Central America and southern Mexico is characterised by a well-defined rainy period from May to October (Hastenrath, 1967; Figure 1). More specifically, however, this rainy season in southern Mexico and most Central American countries is actually characterised by separate maxima in May to July and August to October, with reduced precipitation during the intervening July to August period (Mosiño and García, 1966; Coen, 1973). This so-called "mid-summer drought" (MSD), also known as "veranillo" or “canícula” in Spanish, is characterised by a reduction in precipitation as compared to the wet season averages of up to 40% (Small et al., 2007), although it remains distinct from the actual dry season. A typical annual cycle of precipitation highlights the MSD

signal; climatologically, conditions during the MSD are dominated by drier, warmer, and less cloudier patterns (Hastenrath, 1967), distinct from those seen in the rainy season.

The MSD was identified in the 1960s by Portig (1961), and since then several hypotheses have been proposed to explain the physical mechanisms driving MSDs in the region. Using historical observations as described by Alpert (1945, 1946), Hastenrath (1967) examined the potential for analysis of annual movements of the Intertropical Convergence Zone (ITCZ) and the North Atlantic Subtropical High (NASH) to explain MSDs. This idea was later revisited using a reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hastenrath, 2002), and this identified the key role of the ITCZ in the eastern Pacific with regard to the generation of the MSD. The role of the NASH was also explored in other studies (Giannini et al., 2000; Romero-Centeno et al., 2007).

Magaña et al. (1999) proposed a theory to explain the existence and development of the MSD over the eastern tropical Pacific, which was later revised by Magaña and Caetano (2005) and Herrera et al. (2015). This involved land-ocean-atmosphere processes, and indicated that, in early summer, the warm sea surface temperatures induce convection, corresponding to the first peak of precipitation prior to the MSD. After that, the cloudier skies caused by surface evaporation and enhanced convergence reduce the intensity of the incoming solar radiation, acting to cool the sea surface, reduce convection, and produce a characteristic reduction of precipitation during the MSD. Thereafter, reduced convection leads to increased insolation, allowing the sea surface to become warmed again, resulting in a second peak of precipitation. This theory takes account of the negative feedback associated with regional climate circulation to explain the large range of areas dominated by the MSD. Additionally, this theory also suggests that, if thermal forcing were constant, the rainy season would be persistently dominated by an oscillation formed by peaks and troughs in precipitation, creating a series of MSD-type phenomena (Karnauskas et al., 2013).

Further hypotheses to explain the mechanisms of the MSD from multiple perspectives have been proposed in the last 20 years. The Caribbean low-level jet

has been shown to potentially influence the MSD based on its semi-annual variability, with its two separate maxima in summer and winter (Wang, 2007; Wang and Lee, 2007). Its intensification and regional circulation have also been shown to have implications for the generation of the MSD (Magaña and Caetano, 2005; Herrera et al., 2015). Other potential affecting factors, such as vertical wind shear and atmospheric particles (Angeles et al., 2010), solar declination (Karnauskas et al., 2013) and the Madden-Julian Oscillation (Zhao et al., 2019) have also been explored. It is now generally recognised that the MSD may be triggered by any of a multiple range of processes. The MSD can thus be said to have several physical drivers and contributing factors, and variations in any or all of these might affect the peaks and troughs in annual precipitation across the region.

Current MSD research is challenged by the accuracy of MSD detection and estimation algorithms and the quality and resolution of relevant datasets. Such research has been primarily based on synoptic data (Magaña et al., 1999; Mapes et al., 2005; Small et al., 2007) or monthly climatological data (Curtis, 2002; Karnauskas et al., 2013; Perdigón-Morales et al., 2018). This has left space for a comprehensive assessment of MSD characteristics across space and time and the development of a more consistent definition with broad utility of application by the scientific community to help identify regional driving mechanisms.

This paper applies a regression-based MSD detection algorithm to identify the prevalent atmospheric states over Central America during the MSD. This algorithm was developed based on daily precipitation data and subsequently evaluated using multiple datasets in various formats. Using reanalysis data, the atmospheric states prevalent during various periods of the MSD were determined based on the detected MSD signals from the observed data. A mechanism based on wind-topography interaction is thus proposed to explain the MSD signature over the majority of Central America and Mexico. Using a competitive artificial neural network, the atmospheric state variability during the MSD is estimated by classifying nine pattern types, revealing unresolved variability in the recognised atmospheric drivers of the MSD, which could be potentially classified by considering the El Niño-Southern Oscillation (ENSO) phase. The data and methods used in this research are

introduced in section 2, while section 3 explicitly highlights the observed atmospheric states during MSD periods, including mean states and variability, and potential associations and mechanisms. These results are discussed, and conclusions provided, in section 4.

2 Data and methods

2.1 Precipitation data and other atmospheric variables

The MSD detection scheme utilises multiple daily datasets, including precipitation data from the Tropical Rainfall Measuring Mission (TRMM) (Hong et al., 2007; Huffman 2016), Climate Prediction Center (CPC) global unified gauge-based analysis of daily precipitation (Xie et al., 2007; Chen et al., 2008), and ERA-Interim reanalysis (Dee et al., 2011) in both low (0.5^{o} x 0.5^{o}; ERA/L) and high resolutions (0.25^{o} x 0.25^{o}; ERA/H). Further details of these datasets are shown in Table 1. These datasets were chosen because they:

1) represent various data types, including gauge-satellite blend observations (CPC), satellite blend observations (TRMM), and reanalysis data (ERA), to help develop confidence in the utility and robustness of the algorithm as applied to different datasets; and

2) have been widely used in previous research associated with the MSD over Mexico and Central America (Magaña et al., 1999; Small et al., 2007; Wang, 2007; Karnauskas, 2013), and thus have already been shown to capture the climatology and variability of regional MSD characteristics in this region (e.g. Small et al., 2007; Diro et al., 2012; Karnauskas, 2013).

Other atmospheric variables used in this research include the 2 m air temperature, 10 m wind speeds, surface pressure, and cloud fraction, all of which were extracted from the ERA/L for the period from 1993 to 2017. The sea surface temperature during 1993 to 2017 over the domain was extracted from the NOAA OI SST V2 High Resolution Dataset (Reynolds et al., 2007).

2.2 MSD detection algorithm

As the MSD has a regionally varying signature, the methodology used for the determination of its characteristics largely depends on the objective of a given study. Methods used to detect the MSD can, however, be summarised based on studies to which they are applied and/or their statistical features. Several methods to detect the MSD over Central America and Mexico have been previously proposed. Many researchers focusing on the physical dynamics of the MSD have used spatially-averaged precipitation time series to capture typical MSD characteristics for particular regions (Magaña et al., 1999; Mapes et al., 2005; Small et al., 2007); however, this approach ignores MSD spatial variability and may also fail to reveal certain climatological characteristics seen only on relatively small spatial scales. Additionally, this approach tends to treat all potential MSD events as a composite, rather than examining each as a single event in each year, thus ignoring the interannual variability of the MSD.

Methods used in more recent studies include specially designed algorithms or indices that detect the MSD and quantify its intensity. Karnauskas et al. (2013) used climatological precipitation to detect the global distribution of the MSD and analysed the differences between relative maxima and minima to determine the intensity of these events. This form of detection notably references climatology rather than the MSD in each particular year, however; thus, the application of monthly data limits the accuracy of detection which may cause the MSD signature to become less clear. Perdigón-Morales et al. (2018) used station data to examine the MSD distribution in Mexico, dividing the climatological monthly precipitation time series in Mexico into five types of precipitation with MSD in 1) only July, 2) only August, 3) July to August, and 4) June to August, along with 5) precipitation with an insignificant MSD signal. Two indices were used to quantify the intensity of the detected MSD. These were a separate percent diminished (PD) index to measure the MSD with respect to peak precipitation, and a percent accumulated (PA) index referencing cumulative rainfall during the MSD.

Major issues identified using these MSD detection methods include 1) the determination of the MSD area and the quantification of MSD characteristics not being well separated; 2) the precision and practicability of detected MSDs being

limited by the application of monthly precipitation data; and 3) the fact that, as most algorithms only examine climatological MSD, they ignore the potential importance of the underlying variability in the composites. Additionally, most MSD detection algorithms used previously have been applied to the two peaks in May to July and August to October separately, thus assuming that the reduced precipitation between these two peaks is the critical factor determining an ideal MSD signal. However, algorithms applied in this way run the risk of failing to distinguish multimodal distributions from the required bimodal MSD signals when applied to daily precipitation data due to the existence of anomalous noise (Zhao et al., 2019).

The current work proposes and develops a new algorithm to determine the MSD area, to the extent that it exhibits MSD signals, and to better estimate annual MSD characteristics. The protocol for this algorithm is as follows:

a. Identification of MSD classified locations

A dataset containing daily precipitation P (X, Y, t), where X and Y are east and north grid coordinates and t is time in days, is made available to the algorithm. For each such dataset, the first step is to calculate its annual climatology P_{clim} using the full P record. P_{clim} is calculated as a climatological annual cycle, in Julian days, with data on the non-existent February 29th in each non-leap year filled by the mean of that on February 28th and March 1st. Hence, P_{clim} is always a dataset of size (X, Y, 366). The MSD signal is detected at each (X, Y) coordinate independently, and the climatological precipitation time series P_{clim} is also smoothed using a 15-day window with a Gaussian-weighted moving average at each grid point. The resultant data are recorded as P_{sm}. For the time series P_{sm} (X, Y), the existence of the MSD signal is detected and confirmed based on three criteria:

1) two maximum precipitation peaks, P_{max1} and P_{max2}, should exist separately in the periods May 15th to July 15th and August 15th to October 15th; and their corresponding dates are separately recorded as d1 and d2;

2) d3, which corresponds to the date of annual maximum precipitation, should thus be the same as either d1 or d2; and

3) the linear trend in the precipitation time series between January 1st and d1 should be significantly positive, while the trend between d2 and December 31st should be significantly negative, to indicate the significance of precipitation enhancement/reduction before/after an MSD event. If these three criteria are satisfied, the coordinate (X, Y) is identified as being a representative location for the existence of the MSD; otherwise, the location is deemed to not typically display the characteristics of the MSD signal. The linear trends outlined here are analysed by linear regression. In this algorithm, the existence of two rainfall peaks during rainy seasons ensures the existence of precipitation during the MSD, while the significance of the two linear regressions further guarantees the bimodal shape of annual precipitation. Although there is still the possibility that a third peak of precipitation could exist during a period of detected MSD signals, time series that satisfy the three criteria outlined above are still classified as MSD signals due to the established precipitation reduction and the recognised dominance of the annual bimodal distribution.

b. Detection and quantification of annual MSD signals at classified locations

For MSD (x, y) classified locations, the MSD signal in each year is determined following the procedures as outlined above for individual years rather than being averaged. For each detected MSD signal, several fundamental metrics can thus be determined, including the onset date (the date of P_{max1}), end date (the date of P_{max2}), and duration (length of days between onset and end dates). The intensity of each detected MSD signal is quantified by the intensity of the MSD (I_{msd}) as defined by García-Martínez (2015):

$$I_{msd} = \frac{P_{max}\text{-}P_{min}}{P_{max}}$$

where P_{max} is the larger of P_{max1} and P_{max2}, and P_{min} is the temporal mean of daily precipitation across the MSD period. I_{msd} and its proxies have previously been used to quantify the strength of the MSD signal in Costa Rica (García-Martínez, 2015; Zhao et al., 2019) and Mexico (Perdigón-Morales et al., 2018). In this study, I_{msd} has thus been calculated for both climatological and annual MSD signals.

The algorithm was applied and tested using multiple precipitation datasets (TRMM, CPC, ERA/H, and ERA/L), tasked with examining the frequency, duration, onset and end Julian dates, and climatological I_{msd} as calculated from P_{clim} at each grid location. Once evaluated, this algorithm was applied to the detection of the MSD signals over Mexico and Central America at each grid location, using CPC daily precipitation as input data. Evaluation of the method was undertaken using four datasets (CPC, TRMM, ERA/H and ERA/L); based on the results, the CPC data was then selected to conduct the following studies. The CPC data were constructed based on rain gauge observations of daily precipitation over global land areas from over 30, 000 stations (Xie et al., 2007). The version used here was initialised at 0.125^{o} and released with 0.5^{o} spatial resolution to minimise interpolation errors. Currently, the CPC precipitation data provides the most accurate observations covering both tide gauges and satellites with daily resolution over all tropical and subtropical regions where high gauge network density exists (Xie et al., 2007). The time span of the CPC data covers all periods since 1979, allowing extraction of the time series data from 1993 to 2017 to reflect the behaviour of MSD events over the past 25 years. As the CPC data only include precipitation over land, the study focus is on MSD events over land.

2.3 Self-Organising Maps

Several statistical methods have been proposed to identify important patterns of variability in large spatial-temporal climate datasets. Empirical orthogonal function (EOF) analysis (Lorenz, 1956) offers a variance-based statistical dimensionality reduction technique that has been widely used in synoptic and theoretical climatology. This method separates the total variance of the variables into a set of orthogonal (statistically uncorrelated) modes (eigenvectors) by creating linear combinations of those variables. In this respect, EOF analysis is equivalent to principal component analysis (PCA). Nevertheless, it is important to recognise that EOFs are a mathematical construct, and that eigenvectors (modes) are not necessarily physically interpretable, although the first few modes that explain the highest percentages of the total data variance are often physically meaningful and may well be representative of real climate modes.

Another set of dimensionality reduction methods, cluster algorithms, have been applied to various aspects of synoptic climatology. K-mean clustering (Steinhaus, 1956; Hartigan and Wong, 1979), one of the most frequently used cluster algorithms, is also often used to determine the dominant climatological patterns within oceanic and atmospheric variables. For a dataset with both spatial and temporal dimensions, K-mean clustering aims to identify a set of cluster centroids that minimise the distances between the temporal fields allocated to each cluster centroid while maximising the distances between pairs of cluster centroids. This procedure is mostly achieved by iteration with random selected initial cluster centroids.

In terms of generating clusters, self-organising maps (SOMs) (Kohonen, 1990) achieve similar results to K-mean and other distance-based cluster methods. As with other dimensionality reduction methods, SOMs aim to reduce high-dimensional datasets into datasets with relatively low numbers of dimensions, typically only two (Vesanto et al., 1999); the generated patterns from SOMs are then organised topologically. SOMs tend to place similar clusters statistically into relatively closer locations in low-dimensional maps and to separate distinct clusters. This characteristic makes SOMs powerful visualisation tools where there are distinct characteristic states present in a given dataset. Previously, SOMs have been used to examine ocean current variability on the West Florida Shelf (Liu and Weisberg, 2005), analyse North Atlantic climate variability (Reusch, 2007), detect El Niño-Southern Oscillation flavours (Johnson, 2013), extract patterns from coastal model outputs (Williams et al., 2014), measure the variability of marine heatwaves off eastern Tasmania (Oliver et al., 2018), and determine weather regimes over southern Mexico (Díaz-Esteban, and Raga, 2018). The basic mathematical and statistical details of SOMs can be found in Kohonen (1990).

Here, the SOM approach is applied to analyse regional atmospheric states during each MSD event as detected in the CPC daily precipitation data from 1993 to 2017 over Mexico and Central America, using the algorithm discussed in section 2.2. To achieve this, anomalous 2 m temperatures, 10 m winds, and surface pressures were temporally averaged across each detected MSD event from the start date to the end date in the domain (120^{o}W to 60^{o}W, 0^{o} to 30^{o}N), based on ERA-Interim reanalysis.

Prior to applying the SOM, these data were scaled by the mean and a spatial standard deviation. The variables in each SOM node were thus reconstructed to obtain physically explainable patterns by re-adding the mean and multiplying by the same standard deviation.

Choice of SOM size (map size) is critical for the application of the technique (Gibson et al., 2016). An ideal SOM is expected to be large enough to discriminate between all possible characteristic states while reducing the dimensions of the datasets as much as possible. A suitable map size for SOMs should therefore maximise the similarity within each pattern and minimise the dissimilarity between each pair of patterns (Lourenco et al., 2004). Based on this, a correlation-based method was used to determine the optimal map size for the SOMs. As more nodes were included, the generated patterns were reconstructed into a dataset of the same size as the original data by duplicating each pattern based on its allocated temporal data; the correlations between these two datasets were then calculated. The final map size of the SOM was determined as the size at which the correlation tended to a constant and its first difference tended to be stationary, a procedure illustrated in Figure 2. It was determined that for SOM configurations greater than (3, 3), the increase in correlations tended to be small and the first difference tended to be stationary; thus, (3, 3) was chosen as an applicable map size for SOM in this study. It should be noted that this method of choice introduces some subjectivity; different observers may choose different map sizes. However, since the main use of SOMs in this research is to illustrate the variability of atmospheric states associated with MSD event signatures rather than accurately detecting all potential atmospheric characteristics of the MSD, any small biases are unlikely to significantly affect the conclusions. After obtaining the output of the SOM at (3, 3), the average atmospheric states during onset, peak, and end of the MSD event signatures in each node of the SOM were calculated based on the cluster results of the SOM.

2.4 Atmospheric states during the MSD

In our analysis, the atmospheric states during the MSD over Central America and Mexico were determined by calculating the temporal average of the atmospheric variables in all identified annual MSD event signatures. From the annual MSD

signatures in the CPC data, anomalies in the 2 m temperature, 10 m wind, and surface pressure were temporally averaged across the start, end, and peak (corresponding to the minimum precipitation during an MSD signal) dates for all detected MSD events. The benefits of using this method to determine the mean states during MSD events include ensuring the continuity of the identified state at each grid location without reducing the resolution of the MSD detection. Similar approaches were used by Oliver et al. (2018). The atmospheric states for SOM nodes are determined in the same way, but only for detected signals in the corresponding nodes.

3 Results

3.1 Evaluation of the MSD detection algorithm

The MSD detection algorithm was evaluated by applying the algorithm to four datasets containing daily precipitation measures in the region of Mexico and Central America with the aim of detecting all MSD event signatures at each grid cell location and thus examining the spatial frequencies (Figure 3a, S1a, S2a, S3a), the mean onset (Figure 3b, S1b, S2b, S3b) and end (Figure 3c, S1c, S2c, S3c) dates, and durations (Figure 3d, S1d, S2d, S3d) of the events, as well as the climatological intensity I_{msd}, as calculated by P_{clim} (Figure 3e, S1e, S2e, S3e). Despite the use of very different datasets, the results proved to be remarkably consistent, offering confidence that the technique is robust to data choice and thus has increased utility. According to the results, the bimodal distribution of precipitation is evident across most of Mexico and Central America, with the exception of north-western Mexico and the Caribbean coast of Central America; this includes adjacent oceanic regions including the eastern Pacific warm pool. In most of these regions, the MSD signatures exhibit robust characteristics in terms of annual precipitation, revealed by the high frequency of MSD events (Figure 3a). The spatial variability of the MSD over this domain is also revealed by the algorithm. Towards the southeast, the MSD tends to start earlier and end later, creating a relatively longer duration, and some differences also exist between land and ocean signals, shown in the dramatically long MSDs over the North Atlantic close to Cuba and part of the Caribbean Sea (Figure 3b, c, d). The MSD signatures in central Mexico tend to start in early July and

end in early September, while those in southern Mexico and Central America are generally dominated by longer durations, from June to September and from late May to early October, respectively. It should be noted that the long MSD in Central America seems somewhat inconsistent with that observed in Costa Rica, however: although the first “Canicula” (a few days without rain) normally occurs in early June, the most consistent precipitation reduction tends to occur from mid-July to mid-August, and in some years, this can extend to late August (Brito et al., 2014). This difference could be associated with the definition of the MSD used here, which defines the MSD as the period between two precipitation peaks; thus, using this definition, the MSD defined here contains periods with both precipitation reduction and rainfall peaks. Additionally, the application of daily precipitation data introduces more potential rainfall peaks, subsequently extending the detected periods of MSDs.

The strength of the climatological MSD shows relatively large spatial variability (Figure 3e). Measured in terms of I_{msd} as calculated from P_{clim} at each grid location, intense climatological MSD signatures tend to exist in the oceanic regions around Jamaica, eastern Cuba, and western Bermuda, as well as on the Pacific coast of Mexico and Central America. The performance of detected MSD intensity can vary with data, shown by the fact that the anomalously intense MSDs in the coastal regions along the northern Yucatán Peninsula, could be observed in ERA/L (Figure S2e) and ERA/H (Figure S3e), while it is not shown in other datasets used in this study. Notably, any MSD signature was generally absent along the Caribbean coast of Central America, consistent with findings of previous research (Magaña et al., 1999; Taylor and Alfaro, 2005; Small et al., 2007; Amador, 2008; Zhao et al., 2019).

Although the results from the four datasets were generally consistent, some biases were evident. The detected MSD signals showed relatively high variability in TRMM compared to those seen in the other three datasets, as well as an absence of MSD on the Pacific coast of central Mexico and the Gulf of Mexico. The MSD signatures detected in the oceanic regions around Cuba and the Caribbean Sea were stronger in TRMM than in the ERA datasets. Additionally, robust, intense, and long MSD signals were observed over the Yucatán Peninsula using ERA/L (Figure S2) and ERA/H (Figure S3), with greater variability seen in other datasets for this region. In spite of these

differences, the spatial variability of the MSD signals detected in the different datasets was generally consistent, confirming that the algorithm is robust in terms of capturing the MSD signature across the domain for various data types and resolutions.

3.2 Near-surface states during MSD

Using all annual MSD events identified in the CPC datasets, the mean atmospheric states (anomalies of 2 m temperature, 10 m wind, surface pressure, CPC precipitation, cloud fraction, and sea surface temperature) were calculated for the start, end, and peak dates of MSD event signatures over Central America and Mexico (Figure 4). During the onset and end of the MSD events across the region, remarkably similar atmospheric states emerged. As MSDs begin and end, enhanced westerly wind anomalies approach the Pacific coast of southern Mexico and Central America, while enhanced easterly wind anomalies appear at the coast of Mexico and the nearby Gulf of Mexico, accompanied by positive precipitation anomalies over those regions exhibiting MSD characteristics. These two onshore wind anomalies contribute to an anomalous cyclonic centre in southern Mexico, and during this period, southern Mexico, Central America, and the eastern Pacific warm pool are generally dominated by negative temperature anomalies, which becomes more obvious at the onset of the MSD. During the onset and end of the MSD, a dipole pattern forms between surface pressure anomalies below and at 12^{o}N; similar dipoles can be detected in temperature patterns; however, these are less obvious. At the driest point of the MSDs, opposing atmospheric patterns are exhibited in this region, dominated by reduced precipitation, anticyclonic anomalies in southern Mexico, and offshore winds in the Pacific coast of Central America and coast around the Gulf of Mexico.

To dynamically illustrate the propagation of the atmospheric states during MSDs, the mean states were calculated for five different periods (onset dates, onset dates + 15 days, peak dates, onset dates + 50 days, and end dates, as illustrated in Figure 4). These time points were chosen to highlight the climatological patterns in characteristic periods of MSD generation and to generally illustrate the process of MSD development. When MSDs start over Central America and Mexico, the

atmospheric states in the region are dominated by a cyclonic anomaly centre accompanied by multiple positive precipitation anomalies. At about 15 days after onset, the cyclonic anomalies weaken and the westerly (enhanced eastward) wind anomalies in the Pacific approaching the Pacific coast are replaced by easterly (westward) wind anomalies that are accompanied by reduced precipitation. At the same time, an extension of the NASH occurs over Florida and Cuba. Later, in the period around the peak dates (40 days after onset dates on average), anomalous westerly winds become prominent in the Gulf of Mexico, inducing an anticyclonic anomaly circulation centred in southern Mexico. This anticyclonic system is also contributed to by the westward propagation of the NASH extension determined in the previous period (15 days after onset dates). At the same time, the intensity of negative precipitation anomalies over land increase. After about 10 more days, roughly 50 days after the onset date, the dominant anticyclonic anomaly circulation becomes insignificant, and the easterly wind anomalies from the Caribbean Sea develop into westerlies that head towards the Pacific coast, while precipitation increases. This pattern expands to form a significant cyclonic anomaly system around 90 days after the onset date.

Several notable characteristics can be identified in these results. The wind field over land is generally insignificant compared to that over oceanic regions, except for coastal regions in Central America and Mexico near the Gulf of Mexico. This may be associated with topographic influences, in that high and steep topographies over Central America and Mexico tend to block coming winds and induce topographic uplift (Figure 1a). In addition, constant dipole patterns exist to the north and south of 12^{o}N, especially at the onset, end, and peak of MSD events.

It is also notable that offshore wind anomalies are generally weaker than onshore wind anomalies during the MSD. For example, onshore winds approaching the Pacific coast of Central America are clearly more significant than offshore winds at the Caribbean coast, during the onset of the MSD (Figure 4a) . During the development of bimodal precipitation, the onshore winds uplift as they approach mountains, discharging most of their moisture as precipitation. The major part of these now-dry winds passes over the mountains to become offshore winds on the other side of the

domain, while a relatively small part of the winds is reflected back. This feature induces the phenomena that onshore winds in one side of the domain are generally larger than the offshore ones on the other side, a characteristic discussed by Zhao et al. (2019).

As an important factor influencing the MSD over the Caribbean region and Central America, the Caribbean low-level jet shows some signatures in the presented results, such as the strong easterlies passing from the Caribbean Sea to the tropical Pacific during the development of the MSD (Onset + 15 days to Peak, as seen in Figure 4). During the onset to the peak of the MSD events, the strength of this flow tends to increase, while a decrease occurs during the peak and then to the end of the MSD event. This feature indicates that a stronger Caribbean low-level jet is positively correlated with the precipitation reduction seen during the MSD, corresponding to effects identified in previous work (Amador, 1998; Chelton et al., 2000; Wang, 2002; Gamble et al., 2008).

3.3 Typologies of atmospheric states during MSDs

The mean MSD metrics, including onset, peak, and end dates of MSD, and I_{msd}, are calculated for each SOM node and are shown as differences from the mean for all MSD events (Figure 6). Atmospheric mean states (anomalies of air temperature, wind, and surface pressure) during MSDs are shown in Figure 5 (based on our SOM analysis) along with their mean states during the onset (Figure 7), peak (Figure 8), and end (Figure 9) periods of MSDs in each SOM node. The precipitation (Figures 10 to 12) and sea surface temperature (Figure S4-6) anomalies in each SOM node during onset, peak, and end periods for the MSD are also reconstructed. Each node is denoted as Node (i, j), where i and j range from 1 to 3, resulting in a total of nine node types. With respect to the atmospheric mean states, which are the reconstructed outputs from the SOM, the pattern of typologies is well-organised in a dynamic manner. In cases of increase of i, Node (i, j) tends to exhibit opposing patterns to the original node. Node (3, 2), for instance, is dominated by negative pressure anomalies on the Caribbean side, positive pressure anomalies in the eastern Pacific warm pool, and westerly (eastward) wind anomalies towards the Pacific coast of Central America in remarkable opposition to Node (1, 2). In cases of

increase of j, Node (i, j) tends to exhibit cooler temperature anomalies, higher pressure anomalies, and enhancement of the easterly wind flow. Specifically, for Nodes (2, j) (j=1,2,3), the pressure anomalies tend to increase as j increases, corresponding to enhancement of the easterly wind anomalies. The four nodes with the most extreme atmospheric states during MSD periods are allocated to the corners of the SOM maps.

In terms of the atmospheric states during the onset, end, and peak of MSD events and the associated MSD metrics for each SOM node, many interesting features arise in the generated patterns. Based on the revealed signatures, the nine nodes can be further classified into five groups (G1 to 5) divided by the dominant atmospheric states and MSD characteristics revealed by the SOM.

G1. G1, which includes only Node (1,1), represents the most intense and longest (starting earlier and ending later) MSD event (see Node (1,1) in Figure 6). Precipitation patterns around the Pacific coast of the domain in this group are negative or insignificantly positive during the onset and end of the MSD (see Node (1,1) in Figures 10 and 12), and are generally negative during the peak of the MSD (see Node (1,1) in Figure 11), indicating that MSD events in this group are focussed in that area. During the MSD period, the atmospheric states are consistently dominated by low-pressure systems in the eastern Pacific, insignificant near-surface anticyclone centring in the Gulf of Mexico, and clear wind transports from the Caribbean regions to the Pacific side (see Node (1,1) in Figure 5). The near-surface wind-pressure systems are generally constant across all typical time points (onset, peak, end) of the MSD, with the exception of relatively strong anticyclonic activity in the Gulf of Mexico during the peak and end of the MSD (see Node (1,1) in Figure 8), which indicates the westward propagation of the NASH. Significant air (Node (1,1) in Figure 7, 9) and oceanic (Node (1,1) in Figure S4, S6) near-surface warming around the tropical Pacific exist during the onset and end of the MSD at this node, while a relatively weak signal is detected in the pattern during the peak (Figure 8, S5). This feature agrees with Magaña et al. (1999), which used the vertical convection induced by the warming and cooling of the sea surface temperature to reveal the mechanisms of the MSD.

G2. G2, which includes Node (1,3), Node (2,2), and Node (2,3), represents a group of relatively strong and long MSD events (see corresponding nodes in Figure 6). MSD events in this group happen mainly in southern Mexico, with enhanced rainfall patterns during the onset and end of the MSD (Figure 10, 12) and opposite patterns during the peak of the MSD (Figure 11). The near-surface atmospheric states during MSD events in this group are generally dominated by high pressure systems over the domain, weak convergence, and relatively significant winds from the Caribbean areas to the eastern Pacific, passing through Central America (Figure 5). The MSD events in this group start with onshore wind anomalies over the Pacific coast of Central America and an extension of NASH, which is seen more significantly in Nodes (2,2) and (2,3) (Figure 7). The subtropical high then moves east, bringing enhanced easterlies from the Caribbean Sea, corresponding to the increased strength of the Caribbean low-level jet during July (Figure 8). These easterlies could strengthen the moisture flux divergence, weakening the vertical convection, and finally contributing to the low-level high pressure system and weak anticyclonic system seen over southern Mexico, corresponding to the precipitation reduction observed during the onset to peak phase of the MSD. The major source of internal variability in this group exists in the wind-pressure patterns during the end of MSD (Figure 9), with generally low-level low pressure cyclonic systems in Node (2,2), dipole patterns located in the Gulf of Mexico and North Atlantic in Node (1,3), and generally high pressure systems in Node (2,3). This variability indicates three potential propagation vectors of the westward extension of the NASH after precipitation reduction begins. For Node (2,2), the extension moves northward, inducing relatively high pressure over the southern part of the United States during late summer in corresponding years (Rajagopalan, 2000; Dominguez et al., 2010); for Node (1,3), the energy of the westward extension dissipates and is replaced by a low-pressure system, forming dipole patterns with the original NASH; in Node (2,3), the westward extension of the NASH is relatively prolonged, inducing a generally dry late summer. There is also general cooling over the domain during the onset and end of the MSD in this group, while weak warming is observed around the Pacific coast of southern Mexico during the peak of the MSD, which may be induced by the high-pressure system centred over southern Mexico at that time, indicating that the temperature may play a role in the events seen in this group, albeit weakly.

G3. G3, which includes only Node (2,1), consists of moderately strong and long MSD events (Figure 6). During MSD events in this group, the domain is dominated by generally low pressure systems and offshore wind anomalies from westerlies from the central Pacific, accompanied by weak but consistent positive temperature anomalies over the Caribbean Sea (Figure 5). During the onset of the MSD, a low-level cyclonic system with significant low pressure anomalies exists over the Gulf of Mexico, inducing north-westerlies along the eastern coast of southern Mexico (Figure 7). The westerlies from the Pacific warm pool then bring onshore wind anomalies to the western coast of Central America. The near-surface wind-pressure systems seen during the peak and end of the MSD in this group exhibit similar signatures, being characterised by weak NASH, generally low surface pressure, and weak cyclonic systems contributed to by the onshore winds in the Gulf of Mexico and the Pacific coast of Central America (Figures 8, 9). The MSD events in this group are potentially determined by the precipitation changes brought on by the suppression then enhancement of the low pressure cyclonic system during the onset to peak, then peak to end, phases of the MSD, which appear to weaken or strengthen, respectively, the convergence and vertical convection over the domain.

G4. Consisting only of Node (1,2), G4 represents MSD events of insignificant strength and duration, with earlier onset and end dates (Figure 6). Climatologically, the near-surface atmospheric states during the MSD in G4 are dominated by high surface pressure over the Caribbean Sea and the northern part of the land, with low surface pressure in the eastern Pacific and strong westerlies from the central Pacific (Figure 5). Across the typical time points (onset, peak, and end) of MSD events in this area, one of the most significant features of this group is the onshore wind anomalies that appear separately in the Pacific and Caribbean Seas, thus existing on both coasts of Central America (see Node (1,2) in Figures 7 to 9). During the onset and end of the MSD in this group, these two-sided onshore wind anomalies induce orographic uplift due to the existence of significant mountains across Central America (Figure 7, 9), which enhance the rainfall corresponding to the two peaks during the MSD (Figure 10, 12). During the peak of the MSD in this group, a significant extension of the NASH also induces south-easterly wind anomalies (Figure 8), which could potentially be due to the Caribbean low level jet (Maldonado et al., 2018), seen along the coast of

583 the Gulf of Mexico, transporting water vapor from the Caribbean Sea to eastern
584 Mexico. This feature reduces the amount of moisture arriving at Central America,
585 corresponding to a precipitation reduction compared to that observed during the
586 onset and end of the MSD.

587 ***G5***. G5 consists of the weakest and shortest MSD events (Figure 6), thus including all
588 Nodes in the (3, j) system (j = 1,2,3). During the MSD events in this group, relatively
589 high and low pressure systems exist below the 12^{o}N level, respectively, separated by
590 significant westerlies and north westerlies (Figure 5). During the onset and end of
591 MSDs in this node, low-level low pressure systems and associated cyclonic
592 circulation bring onshore wind anomalies to the coast of Central America and the
593 Gulf of Mexico, enhancing the precipitation in these regions (Figure 7, 9). The
594 geostrophic balance of strong westerlies from the Pacific indicates the existence of a
595 pressure gradient across the 12^{o}N line. During the peak of MSDs in this group, the
596 low pressure cyclonic system moves eastward to centre itself in the Gulf of Mexico
597 or the North Atlantic (Figure 8), removing some of the moisture from this area to the
598 Caribbean Sea. This feature partially induces the precipitation reduction seen during
599 the MSDs in this node, although westerlies from the Pacific remain and contribute to
600 rainfall generation, making the MSDs in this group relatively weak with shallow
601 precipitation troughs.

602 Despite various features appearing across the SOM typologies, some general
603 signatures of MSD characteristics can be summarised as follows:

604 1) Consistent easterlies from the Caribbean Sea to Central America throughout
605 all relatively strong MSD events, as included in G1 to 4 (with the exception of Node
606 (2,1)), accompanied by relatively high pressure systems in the northern domain and
607 opposing pressure patterns in the southern domain, which are significant in Nodes
608 (1, j) (j=1,2,3), but less significant in Nodes (2, j) (j=1,2,3). In contrast, weak MSD
609 events are dominated by contrasts (G5; Nodes (3, j) (j=1,2,3)). Here, changes
610 between easterlies and westerlies suppress and enhance the orographic
611 precipitation over the Pacific side of Central America, respectively, inducing generally
612 drier or wetter summer periods and subsequently stronger and weaker MSD events.

This feature is highly consistent with the mechanisms proposed in Section 2, as observed in previous studies (Magana et al., 1999; Small et al., 2007; Amador, 2008).

2) The development of relatively strong MSDs (G1 to 4 and Node (3,3) in G5) is highly influenced by the NASH. The westward propagation and further extension of this to the continental high-pressure system over southern Mexico reduces levels of moisture transport from the Caribbean Sea to Central America and adjacent areas, contributing to the precipitation change seen during the MSD. Easterlies originating from Caribbean areas also contribute to this process, which can thus be associated with the peak in the Caribbean low-level jet in summer (Small et al., 2007; Hidalgo et al., 2015; Martinez et al., 2019).

3) The general dominance of onshore wind anomalies during the onset and end of the MSD, and the prevalence of offshore wind anomalies during the driest point of each MSD (MSD peaks) is no longer significant when shown in terms of atmospheric mean states during MSD events. Additionally, any consistency between onset and end patterns, and the contrast between onset/end and peak patterns, is not expressed in the SOM nodes.

To explain this variability, Figure 13 illustrates the MSD occurrence time series by plotting the proportion of MSDs detected in each SOM node. The colours in Figure 11 indicate whether at least four months in the period from September to February can be identified as an ENSO event (El Niño or La Niña phase), based on data adopted from NOAA CPC. If so, the year corresponding to September to December in the period during September to next year's February is identified as an ENSO year. The NOAA CPC specifies an ENSO episode when the three-month running mean Niño-3.4 region sea surface temperature anomaly in the domain [5ºN-5ºS, 170ºW-120ºW] is greater than 0.5ºC (El Niño) or less than -0.5ºC (La Niña) for at least five consecutive three-month periods. Figure 11 reveals that MSDs evident in certain years thus dominate most SOM nodes: for instance, 83.13% of MSDs in Node (1, 2) existed in 2016 and 41.71% of MSDs in Node (2, 2) existed in 2013. This characteristic causes the time series of SOM nodes to become low-frequency signals with less variability, significantly distinct from those of total MSDs (see 'Total' in Figure 13). Thus, each SOM node only displays variability for MSDs over a relatively short time

scale (several years) rather than the total variability from 1993 to 2017. A working hypothesis for this is that the absence of consistency at onset and end, and contrast between peak and onset/end in the atmospheric mean states of MSDs in the SOM nodes indicates that consistency and contrast during MSDs can only be detected over relatively long time-scales.

To quantify the statistical connections between occurrences of MSDs and ENSO events, the proportion of MSDs in three specific ENSO periods (El Niño years, La Niña years, and neutral years) was calculated for both the total number of MSDs and for each SOM node (Table 2). Statistical significance at the 90% level for each SOM node was determined by using a bootstrap approach with 1,000x duplication. The tendency for MSDs to occur in particular ENSO years was not found to be statistically significant, as suggested by the similarity between “Year” and “Total” in Table 3. Nevertheless, the atmospheric states of the MSD signatures in the SOM nodes do show some variability driven by ENSO events. Nearly all nodes tend to be dominated by MSD signatures in a particular ENSO phase: a large proportion of the MSDs in Node (1,1) exist in El Niño years, for example, indicating that ENSO could be a classifier for atmospheric states during MSDs in Central America and Mexico, and thus may modulate MSDs by influencing atmospheric states. The teleconnections between ENSO events and MSD SOM nodes are shown by the opposing influences of positive and negative ENSO phases. Relatively intense MSD events tend to happen in El Niño or neutral years (Nodes (i,j) (i=1,2, j=1,2,3)), while, generally, weaker ones are more prevalent in La Niña years. Positive ENSO phases could produce higher pressure systems in the North Pacific and enhance the NASHs in the boreal summer, inducing easterlies with generally geostrophic balance, corresponding to those revealed in the near-surface pressure patterns of these SOM nodes. Additionally, positive ENSO phases produce positive sea surface temperature anomalies along the Central American coasts (see Node (1,1) in Figures 7 to 10), inducing low atmospheric pressure over the sea and a pressure gradient between the Caribbean Sea and the Pacific Coast. This enhances the easterly winds and favours dry atmospheric conditions over the Pacific Coast and Pacific slope of Central America, particularly where there are gaps in the mountains such as the north side of Costa Rica.

4 Discussion

This paper has proposed a new MSD detection algorithm that utilises daily precipitation data to more accurately characterise MSD signatures and their spatial and temporal variability. The advantages of the outlined approach are 1) the unprecedented use of both climatological and annual precipitation to examine MSD characteristics in particular areas, allowing separate identification and characterisation of MSD regional and temporal signatures at all grid locations, and thus providing richer information about the mean states and variability of MSDs; and 2) the application of daily precipitation information rather than monthly data, as typically used in existing MSD detection methods, which greatly increases the temporal resolution of MSD detection and reduces the likelihood of aliasing the MSD signatures, thus providing greater potential for the exploration of MSD variability over different time scales.

As daily precipitation data are inherently noisy, which is problematic for detection of the MSD signal above such noise, the proposed algorithm uses two additional linear regressions to better estimate and identify the bimodal distribution of the characteristic MSD event annual precipitation signatures. The algorithm was also evaluated using four different datasets, covering multiple data types; it was found that there was remarkable consistency of the results, including similar patterns with results from previous research (Karnauskas et al., 2013; Perdigón-Morales et al., 2018).

The oceanic regions off the Caribbean coast in Central America have been treated as an MSD event region in some prior research (Karnauska et al. 2013); however, a distinct absence of MSD signatures was found in this region within these analyses, mostly due to the large proportion of secondary peaks occurring outside of the period from August 15th to October 15th, the assumed period for the second peak of MSD within the algorithm. The study followed the MSD definition in the recent literature, with the minimum precipitation falling between the two separate peaks identified in May to July and August to October (García-Martínez, 2015), instead of seeking a critical annual bimodal distribution (Karnauskas et al., 2013). However, the algorithm could be easily modified to identify other such signatures by finding peaks

of precipitation with different temporal gaps, rather than using these pre-defined periods.

The characteristics of MSDs over Central America and Mexico were thus identified in this study. Climatologically, MSDs display robust characteristics over a large proportion of this region, though with earlier onsets, later ends, and correspondingly longer durations towards the southeast. It is also notable that stronger and longer MSDs exist off the Pacific coast of Central America, where agriculture represents a large proportion of the economic activity (Gallai et al., 2009). This high spatial variability of MSDs over the domain is consistent with previous studies (Gamble et al., 2008; Gamble and Curtis, 2008; Karnauskas et al., 2013), although here it is shown at a higher resolution.

The physical drivers of the MSD over the domain revealed in this study can be summarised as several separate mechanisms:

- **Large-scale Dynamics**: Generally, the large-scale dynamics during the MSD are dominated by low-level atmospheric patterns. During the onset of the MSD, geostrophically balanced westerlies approach the Pacific coast of the domain, inducing a dipole pressure pattern below and upon 12^{o}N. Some of this system passes through the mountains over Central America and transforms into the south-easterlies that transport moisture from the Caribbean Sea to the Gulf of Mexico and the associated coast. These flows, together with the onshore wind around the Gulf of Mexico, contribute to a low-level cyclonic system centred in the Gulf of Mexico, and this low-pressure cyclonic system covers most of southern Mexico and part of Central America, being accompanied by both land and oceanic cooling. This convergence induces strong upward convection, contributing to enhancement of cloud quantity over the corresponding region, and subsequently inducing an increase in rainfall, corresponding to the onset of the MSD.

  After the onset of the MSD, the westward propagation of the NASH offsets this anticyclonic system, bringing strong easterlies from the Caribbean Sea to the domain and the tropical Pacific. These easterlies, corresponding to

enhancement of the Caribbean low-level jet seen during the boreal summer (Wang, 2002; Hidalgo et al., 2015), may induce significant moisture flux divergence over the Caribbean regions, reducing vertical convection and subsequently inducing a reduction in rainfall. When the westward extension of the NASH approaches the Gulf of Mexico, it induces a low-level divergence zone, which further suppresses vertical convection, reduces rainfall, and induces continental and oceanic heating around the domain. This corresponds to the peak of the MSD. The strength of the anticyclonic system decreases later, weakening the easterlies to the Pacific and subsequently inducing dominance of the westerly trade winds. These westerlies from the tropical Pacific, together with easterlies induced by the NASH, generate a low-level convergence zone, resulting in strong vertical motion, increased cloud quantity, and increased precipitation, creating a second peak of precipitation.

These large-scale dynamics reveal the importance of low-level climate patterns for the generation of the MSD, converging well with the mechanisms proposed in previous studies. As with Magaña et al. (1999), Small et al. (2007) and Martinez et al. (2019), this study highlights the association between the development and decay of the MSD and the reversals in the trade winds. Other factors, such as westward propagation of the NASH and the Caribbean-low level jet, are also fully or partially revealed in the results presented here.

- **Regional forcing**: Regional forcing is another important contributor to the generation of the MSD. It is notable that the southern part of Central America, particularly Panama and Costa Rica, is not covered by the convergence/divergence zone during the generation and development of the MSDs. The MSD in these regions is therefore most likely to be influenced by local factors. During MSD onset, anomalous westerlies and easterlies exist separately in the Pacific off the coast of Central America and the Gulf of Mexico. Influenced by the steep topography that is widely distributed across this region, these wind anomalies tend to move upward (orographic uplift), resulting in enhanced rainfall on the windward coast (orographic precipitation) corresponding to the onset of the MSD in a manner consistent with the link

between the Madden-Julian Oscillation and MSD (Zhao et al., 2019). When easterlies are enhanced during MSD events, this brings precipitation over the Caribbean Sea and dry conditions over the Pacific, with most of the atmospheric humidity precipitated over the mountains as orographic rain**.** When the easterlies weaken, precipitation moves over the Pacific side as low pressures from the west Pacific approach the Pacific coast due to the weak pressure gradient between the Caribbean Sea and the Pacific Coast of Central America, indicating the end of the MSD, particularly in Panama and Costa Rica.

This regional forcing demonstrates the important contribution of topography to the characteristic bimodal annual rainfall signature. Orographic uplift (forcing) of moisture-laden air by onshore winds and the reverse pattern play an important role in rainfall variability, contributing to the characteristics of the MSD. Recognition of the important influence of topography on the MSD over Central America and Mexico is likely to be beneficial to further research on MSDs in other areas. Such bimodal distributions of annual precipitation have been climatologically identified in southern Japan, the east coast of China, the coast of the Gulf of Guinea, and coastal Australia (Karnasuskas et al., 2013), all areas characterised by significant topography; the current research could thus potentially contribute to an understanding of MSD signatures in these areas as well.

Combing the large-scale dynamic and regional forcing to explain the mechanisms of the MSD also highlight the spatial variability of physical drivers across the domain under investigation. These physical drivers, including varying low-level atmospheric patterns, contiguous prolonged climate modes such as the North Atlantic Oscillation, dominant wind flows, and regional signatures, induce varying performance and spatial variability in the MSD across the domain. Individual factors may also contribute to the formation of bimodal annual precipitation over parts of the domain; however, these fail to explain total spatial variability.

This study used an artificial neural network approach to classify the atmospheric states during MSDs over Central America and Mexico into nine patterns (self-organising maps). In the atmospheric states identified, the otherwise clear

similarities between MSD onsets and ends and contrasts between MSD centres and onsets/ends were generally absent. Further, the shift between onshore and offshore wind anomalies during the development of MSD was not as significant as shown by examination of the atmospheric mean states during MSDs. The mechanisms underpinning the MSD explain the existence and development of climatological MSDs over longer time periods (here, 25 years); however, short-term unexplained variabilities among the atmospheric states of MSDs across this region appear to be most highlighted by the varying atmospheric states in the SOM nodes.

The MSD is thus recognised as a regionally significant feature forced by several physical drivers and contributing factors. Previous studies have proposed various mechanisms to explain the signature of the bimodal precipitation, and many of these are fully or partially revealed in the typology of the SOM nodes. These mechanisms include 1) the consistent easterlies during strong MSD events, observed in Nodes (i,j) (i=1,2,j=1,2,3) (Douglas, 1995; Amador and Magaña, 1999; Magaña et al., 1999; Anderson et al., 2000; Small et al., 2007; Gamble and Curtis, 2008; Gamble et al., 2008; Herrera et al., 2015; Martinez et al., 2019); 2) the northward propagation of the ITCZ, indicated by the extension of a low pressure system to the northern part of the eastern Pacific and the near surface convergence during the peak of MSD events (significant in Nodes (2,1) and (2,2)) (Ramírez, 1983; Magana et al., 1999; Curtis, 2002; Hastenrath, 2002; Small et al., 2007; Martinez et al., 2019); 3) the westward extension of the NASH, obvious in various Nodes (Hastenrath, 1976; Taylor et al., 2002; Small et al., 2007; Gamble et al., 2008; Hidalgo et al., 2015; Maldonado et al., 2017; Martinez et al., 2019); and 4) the influence of sea-air-coast interactions motivated by the warming and cooling of the surface sea around the Pacific coast of Central America and Mexico (significant in Node (1,1)) (Magaña et al., 1999; Herrera et al., 2015). These indicate that multiple or individual factors may induce different types of MSD events, suggesting that MSD events with distinct features could be forced by different mechanisms and making their development difficult to summarise in a single theory.

A comparison between this study and results presented in Díaz-Esteban and Raga (2018; DR18 hereafter) is finally made here since both studies used SOMs to analyse

the signature of the MSD, as well as its connection to ENSO. DR18 applied the SOM technique with map size (4, 4) (16 nodes) to daily precipitation patterns during May to October in a 17–year record of the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset, and subsequently generated 16 weather regimes, each of which corresponds to a series of time points. The SOM analysis applied in DR18, therefore, treated data in time point (day) as a single element, which was different from that used in this study, treating composites of climate variables during each MSD event as a single element. DR18 identified several nodes representing MSD periods characterized by a strong signature of the NASH and easterlies over the Caribbean basin. DR18 also clarified that nodes associated with MSD patterns tended to be more frequent in El Niño years, and the ENSO modulation was also shown by significant correlation between ENSO signals and rainfall anomalies. The results in DR18 are generally consistent with those presented here in this study. DR18, however, did not specifically determine the bimodal shape of the annual precipitation during the MSD, meaning that the composites of climate properties in typical MSD time points (e.g. onset, end, and peak dates) could not be explicitly quantified. Further, by only considering MSD events as nodes with precipitation reduction during July and August, various signatures of the MSD (e.g. durations and intensity in this study) were unable to be quantified, with some climate signatures during relatively rare MSD events possibly going undetected, such as the westerlies over the Caribbean coast of Central America during MSD events corresponding to Nodes (i, j) (i=3, j=1, 2, 3) (Figure 5). While DR18 successfully capture the generation and development of the MSD by quantifying the transition among nodes representing different periods of MSD events, here we were able to present a more complete visualization of the major climate patterns during typical time points of MSD events in each SOM node and groups, with greater detail of the spatial variabilities and local signatures.

ENSO influences on MSD events over the domain remain unclear, however. While some studies have suggested that the MSD signal may be weaker during El Niño years (Magaña et al., 1999, 2003; Peralta-Hernández et al., 2008), others have argued that these positive ENSO phases strengthen the tendency to MSD occurrence (Anthony Chen and Taylor, 2002; Curtis, 2002; Hidalgo et al., 2017). Our results

suggest that ENSO could be a potential classifier for the relevant atmospheric states, based on the tendency for MSD occurrence in particular ENSO phases in the examined nodes. Specifically, ENSO's potential to influence MSDs over Central America and Mexico is most likely realised through modulation of the atmospheric states, which have themselves been shown to be contributors to the generation and maintenance of MSDs. As revealed in previous studies (e.g. Jury et al., 2007), these results suggest that positive and negative ENSO phases tend to enhance or suppress, respectively, the intensity of the MSD by inducing or strengthening various atmospheric features, such as the NASH, that are associated with the generation and development of MSDs – and provides us with some optimism for the identification of potential seasonal predictability of MSDs over Central America and Mexico.

References

Alpert, L., 1945. The intertropical convergence zone of the eastern Pacific region (I). *Bulletin of the American Meteorological Society*, *26*(10), pp.426-432.

Alpert, AC, L., 1946. The Intertropical Convergence Zone of the Eastern Pacific Region (II). *Bulletin of the American Meteorological Society*, *27*(1), pp.15-29.

Amador, J.A. and Magana, V., 1999, January. Dynamics of the low level jet over the Caribbean Sea. In *Preprints, Third Conference on Hurricanes and Tropical Meteorology*.

Amador, J.A., 1998. A climatic feature of the tropical Americas: The trade wind easterly jet. *Top. Meteor. Oceanogr*, *5*(2), pp.1-13.

Amador, J.A., 2008. The intra-Americas sea low-level jet: Overview and future research. *Annals of the New York Academy of Sciences*, *1146*(1), pp.153-188.

Anderson, B.T., Roads, J.O., Chen, S.C. and Juang, H.M.H., 2000. Regional simulation of the low-level monsoon winds over the Gulf of California and southwestern United States. *Journal of Geophysical Research: Atmospheres*, *105*(D14), pp.17955-17969.

Angeles, M.E., González, J.E., Ramírez-Beltrán, N.D., Tepley, C.A. and Comarazamy, D.E., 2010. Origins of the Caribbean rainfall bimodal behavior. *Journal of Geophysical Research: Atmospheres*, *115*(D11).

Anthony Chen, A. and Taylor, M.A., 2002. Investigating the link between early season

892 Caribbean rainfall and the El Niño+ 1 year. *International Journal of Climatology:*
893 *A Journal of the Royal Meteorological Society*, *22*(1), pp.87-106.
894 Brito, D.P., León, E.S. and España, W.S. (2014) La Oscilación atmosférica Madden-
895 Julian (MJO) y las lluvias en Costa Rica. Meteorológía y oceanografía, 58.
896 Brunkard, J.M., Cifuentes, E. and Rothenberg, S.J., 2008. Assessing the roles of
897 temperature, precipitation, and ENSO in dengue re-emergence on the Texas-
898 Mexico border region. *Salud pública de México*, *50*, pp.227-234.
899 Chelton, D.B., Freilich, M.H. and Esbensen, S.K., 2000. Satellite observations of the
900 wind jets off the Pacific coast of Central America. Part II: Regional relationships
901 and dynamical considerations. *Monthly Weather Review*, *128*(7), pp.2019-2043.
902 Chen, M., Shi, W., Xie, P., Silva, V.B., Kousky, V.E., Wayne Higgins, R. and Janowiak,
903 J.E., 2008. Assessing objective techniques for gauge-based analyses of global
904 daily precipitation. *Journal of Geophysical Research: Atmospheres*, *113*(D4).
905 Chowell, G. and Sanchez, F., 2006. Climate-based descriptive models of dengue
906 fever: the 2002 epidemic in Colima, Mexico. *Journal of environmental*
907 *health*, *68*(10).
908 Coen, E., 1973. El floklore costarricense relativo al clima. *Revista de la Universidad de*
909 *Costa Rica*.
910 Díaz-Esteban, Y. and Raga, G.B., 2018. Weather regimes associated with summer
911 rainfall variability over southern Mexico. *International Journal of*
912 *Climatology*, *38*(1), pp.169-186.
913 Dominguez, F., Cañon, J. and Valdes, J., 2010. IPCC-AR4 climate simulations for the
914 Southwestern US: the importance of future ENSO projections. *Climatic*
915 *Change*, *99*(3-4), pp.499-514.
916 Douglas, M.W., 1995. The summertime low-level jet over the Gulf of
917 California. *Monthly Weather Review*, *123*(8), pp.2334-2347.
918 Fuchs, A. and Wolff, H., 2011. Concept and unintended consequences of weather
919 index insurance: the case of Mexico. *American Journal of Agricultural*
920 *Economics*, *93*(2), pp.505-511.
921
922 Gallai, N., Salles, J.M., Settele, J. and Vaissière, B.E., 2009. Economic valuation of the
923 vulnerability of world agriculture confronted with pollinator decline. *Ecological*
924 *economics*, *68*(3), pp.810-821.

925 Gamble, D.W., Parnell, D.B. and Curtis, S., 2008. Spatial variability of the Caribbean
926 mid-summer drought and relation to north Atlantic high
927 circulation. *International Journal of Climatology: A Journal of the Royal*
928 *Meteorological Society*, *28*(3), pp.343 - 350.
929 García-Martínez, I.M. (2015) Variabilidad océano-atmósfera asociada a la sequía
930 intraestival en el reanálisis CFSR. MSc thesis, Baja California, Centro de
931 Investigación Científica y de Educación Superior de Ensenada, 71 pp.
932 Giannini, A., Kushnir, Y. and Cane, M.A., 2000. Interannual variability of Caribbean
933 rainfall, ENSO, and the Atlantic Ocean. *Journal of Climate*, *13*(2), pp.297-311.
934 Gibson, P.B., Perkins-Kirkpatrick, S.E. and Renwick, J.A., 2016. Projected changes in
935 synoptic weather patterns over New Zealand examined through self-organizing
936 maps. *International Journal of Climatology*, *36*(12), pp.3934-3948.
937 Hartigan, J.A. and Wong, M.A., 1979. Algorithm AS 136: A k-means clustering
938 algorithm. *Journal of the Royal Statistical Society. Series C (Applied*
939 *Statistics)*, *28*(1), pp.100-108.
940 Hastenrath, S., 1967. Rainfall distribution and regime in Central America. *Archiv für*
941 *Meteorologie, Geophysik und Bioklimatologie, Serie B*, *15*(3), pp.201-241.
942 Hastenrath, S., 1976. Variations in low-latitude circulation and extreme climatic
943 events in the tropical Americas. *Journal of the Atmospheric Sciences*, *33*(2),
944 pp.202-215.
945 Hastenrath, S., 2002. The intertropical convergence zone of the eastern Pacific
946 revisited. *International Journal of Climatology*, *22*(3), pp.347-356.
947 Herrera, E., Magaña, V. and Caetano, E., 2015. Air–sea interactions and dynamical
948 processes associated with the midsummer drought. *International Journal of*
949 *Climatology*, *35*(7), pp.1569-1578.
950 Hidalgo, H.G., Alfaro, E.J. and Quesada-Montano, B., 2017. Observed (1970–1999)
951 climate variability in Central America using a high-resolution meteorological
952 dataset with implication to climate change studies. *Climatic Change*, *141*(1),
953 pp.13-28.
954 Hidalgo, H.G., Durán-Quesada, A.M., Amador, J.A. and Alfaro, E.J., 2015. The
955 Caribbean low-level jet, the inter-tropical convergence zone and precipitation
956 patterns in the intra-Americas sea: A proposed dynamical
957 mechanism. *Geografiska Annaler: Series A, Physical Geography*, *97*(1), pp.41-59.

Hong, Y., Adler, R.F., Hossain, F., Curtis, S. and Huffman, G.J., 2007. A first approach to global runoff simulation using satellite rainfall estimation. *Water Resources Research*, *43*(8).

Huffman, G.J., 2016. The transition in multi-satellite products from TRMM to GPM (TMPA to IMERG). *Version 161025*, p.5.

Johnson, N.C., 2013. How many ENSO flavors can we distinguish? *Journal of Climate*, *26*(13), pp.4816-4827.

Jury, M., Malmgren, B.A. and Winter, A., 2007. Subregional precipitation climate of the Caribbean and relationships with ENSO and NAO. *Journal of Geophysical Research: Atmospheres*, *112*(D16).

Karnauskas, K.B., Seager, R., Giannini, A. and Busalacchi, A.J., 2013. A simple mechanism for the climatological midsummer drought along the Pacific coast of Central America. *Atmósfera*, *26*(2), pp.261-281.

Kohonen, T., 1990. The self-organizing map. *Proceedings of the IEEE*, *78*(9), pp.1464-1480.

Liu, Y. and Weisberg, R.H., 2005. Patterns of ocean current variability on the West Florida Shelf using the self-organizing map. *Journal of Geophysical Research: Oceans*, *110*(C6).

Lorenz, E.N., 1956. Empirical orthogonal functions and statistical weather prediction.

Lourenço, F., Lobo, V., Bação, F., 2004. Binary-based Similarity Measures for Categorical Data and Their Application in Self-Organizing Maps, Internal Report. Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa

Magaña, V. and Caetano, E., 2005. Temporal evolution of summer convective activity over the Americas warm pools. *Geophysical Research Letters*, *32*(2).

Magaña, V., Amador, J.A. and Medina, S., 1999. The midsummer drought over Mexico and Central America. *Journal of Climate*, *12*(6), pp.1577-1588.

Magaña, V.O., Vázquez, J.L., Pérez, J.L. and Pérez, J.B., 2003. Impact of El Niño on precipitation in Mexico. *Geofísica internacional*, *42*(3), pp.313-330.

Maldonado, T., Alfaro, E., Rutgersson, A. and Amador, J.A., 2017. The early rainy season in Central America: the role of the tropical North Atlantic SSTs. *International Journal of Climatology*, *37*(9), pp.3731-3742.

Maldonado, T., Alfaro, E.J. and Hidalgo, H.G., 2018. A review of the main drivers and

variability of Central America's Climate and seasonal forecast systems. *Revista de Biología Tropical*, *66*(1-1), pp.S153-S175.

Mapes, B.E., Liu, P. and Buenning, N., 2005. Indian monsoon onset and the Americas midsummer drought: Out-of-equilibrium responses to smooth seasonal forcing. *Journal of climate*, *18*(7), pp.1109-1115.

Martinez, C., Goddard, L., Kushnir, Y. and Ting, M., 2019. Seasonal climatology and dynamical mechanisms of rainfall in the Caribbean. *Climate Dynamics*, *53*(1-2), pp.825-846.

Mosiño, P. and García, E., 1966. The midsummer droughts in Mexico. In *Proc. Regional Latin American Conf* (Vol. 3, pp. 500-516).

Oliver, E.C., Lago, V., Hobday, A.J., Holbrook, N.J., Ling, S.D. and Mundy, C.N., 2018. Marine heatwaves off eastern Tasmania: Trends, interannual variability, and predictability. *Progress in Oceanography*, *161*, pp.116-130.

Parmenter, R.R., Yadav, E.P., Parmenter, C.A., Ettestad, P. and Gage, K.L., 1999. Incidence of plague associated with increased winter-spring precipitation in New Mexico. *The American journal of tropical medicine and hygiene*, *61*(5), pp.814-821.

Peralta-Hernández, A.R., Barba-Martínez, L.R., Magaña-Rueda, V.O., Matthias, A.D. and Luna-Ruíz, J.J., 2008. Temporal and spatial behavior of temperature and precipitation during the canícula (midsummer drought) under El Niño conditions in central México. *Atmósfera*, *21*(3), pp.265-280.

Perdigón-Morales, J., Romero-Centeno, R., Pérez, P.O. and Barrett, B.S., 2018. The midsummer drought in Mexico: perspectives on duration and intensity from the CHIRPS precipitation database. *International Journal of Climatology*, *38*(5), pp.2174-2186.

Portig, W.H., 1961. Some climatological data of Salvador, Central America. *Weather*, *16*(4), pp.103-112.

Rajagopalan, B., Cook, E., Lall, U. and Ray, B.K., 2000. Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. *Journal of Climate*, *13*(24), pp.4244-4255.

Ramírez, P., 1983. *Estudio meteorológico de los veranillos en Costa Rica* (No. P40/2161). sl](Costa Rica).

1023 Reusch, D.B., Alley, R.B. and Hewitson, B.C., 2007. North Atlantic climate variability
1024 from a self-organizing map perspective. *Journal of Geophysical Research:*
1025 *Atmospheres*, *112*(D2).
1026 Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S. and Schlax, M.G., 2007.
1027 Daily high-resolution-blended analyses for sea surface temperature. *Journal of*
1028 *Climate*, *20*(22), pp.5473-5496.
1029 Rogé, P., Friedman, A.R., Astier, M. and Altieri, M.A., 2014. Farmer strategies for
1030 dealing with climatic variability: a case study from the Mixteca Alta region of
1031 Oaxaca, Mexico. *Agroecology and Sustainable Food Systems*, *38*(7), pp.786-811.
1032 Romero-Centeno, R., Zavala-Hidalgo, J. and Raga, G.B., 2007. Midsummer gap winds
1033 and low-level circulation over the eastern tropical Pacific. *Journal of*
1034 *Climate*, *20*(15), pp.3768-3784.
1035 Small, R.J.O., De Szoeke, S.P. and Xie, S.P., 2007. The Central American midsummer
1036 drought: regional aspects and large-scale forcing. *Journal of Climate*, *20*(19),
1037 pp.4853-4873.
1038 Steinhaus, H., 1956. Sur la division des corp materiels en parties. *Bull. Acad. Polon.*
1039 *Sci*, *1*(804), p.801.
1040 Taylor, M.A., Enfield, D.B. and Chen, A.A., 2002. Influence of the tropical Atlantic
1041 versus the tropical Pacific on Caribbean rainfall. *Journal of Geophysical Research:*
1042 *Oceans*, *107*(C9), pp.10-1.
1043 Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankangas, J., 1999, November. Self-
1044 organizing map in Matlab: the SOM Toolbox. In *Proceedings of the Matlab DSP*
1045 *conference* (Vol. 99, pp. 16-17).
1046 Wang, C. and Lee, S.K., 2007. Atlantic warm pool, Caribbean low-level jet, and their
1047 potential impact on Atlantic hurricanes. *Geophysical research letters*, *34*(2).
1048 Wang, C., 2007. Variability of the Caribbean low-level jet and its relations to
1049 climate. *Climate Dynamics*, *29*(4), pp.411-422.
1050 Williams, R.N., de Souza Jr, P.A. and Jones, E.M., 2014. Analysing coastal ocean
1051 model outputs using competitive-learning pattern recognition
1052 techniques. *Environmental modelling & software*, *57*, pp.165-176.
1053 Xie, P., Chen, M. and Shi, W., 2010, January. CPC unified gauge-based analysis of
1054 global daily precipitation. In *Preprints, 24th Conf. on Hydrology, Atlanta, GA,*
1055 *Amer. Meteor. Soc* (Vol. 2).

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y. and Liu, C., 2007. A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology*, *8*(3), pp.607-626.

Zhao, Z., Oliver, E.C.J., Ballestero, D., Vargas-Hernandez, J.M. and Holbrook, N.J., 2019. Influence of the Madden–Julian oscillation on Costa Rican mid-summer drought timing. *International Journal of Climatology*, *39*, pp.292-301.

ACKNOWLEDGMENTS

This research was supported by a vacation scholarship provided by the Australian Research Council Centre of Excellence for Climate System Science (CE110001028). We also thank Jiale Lou, Yueyang Lu and Yue Man for thoughtful discussions. NJH acknowledges funding support from the ARC Centre of Excellence for Climate Extremes (CE170100023).

FIGURES

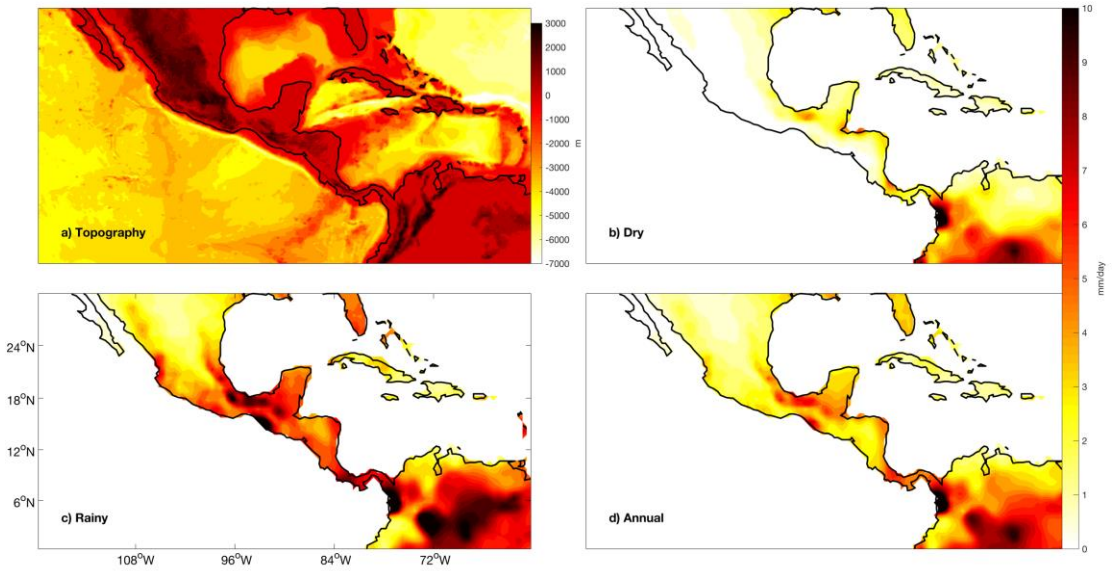

Figure 1. (a) Topography in [120 – 60°W, 0 – 30°N] and average precipitation in (b) dry seasons (January – April, November and December), (c) rainy seasons (May – October) and (d) full seasons. Precipitation used to calculate (b), (c) and (d) is extracted from CPC data during 1993 to 2017.

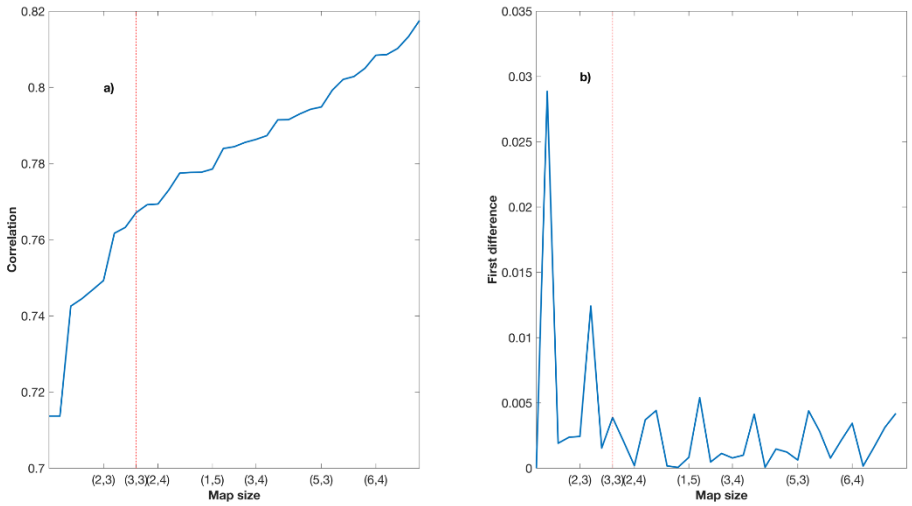

Figure 2. (a) Correlations and (b) associated first difference in different SOM map sizes. Determined SOM map size is indicated by red dashed lines.

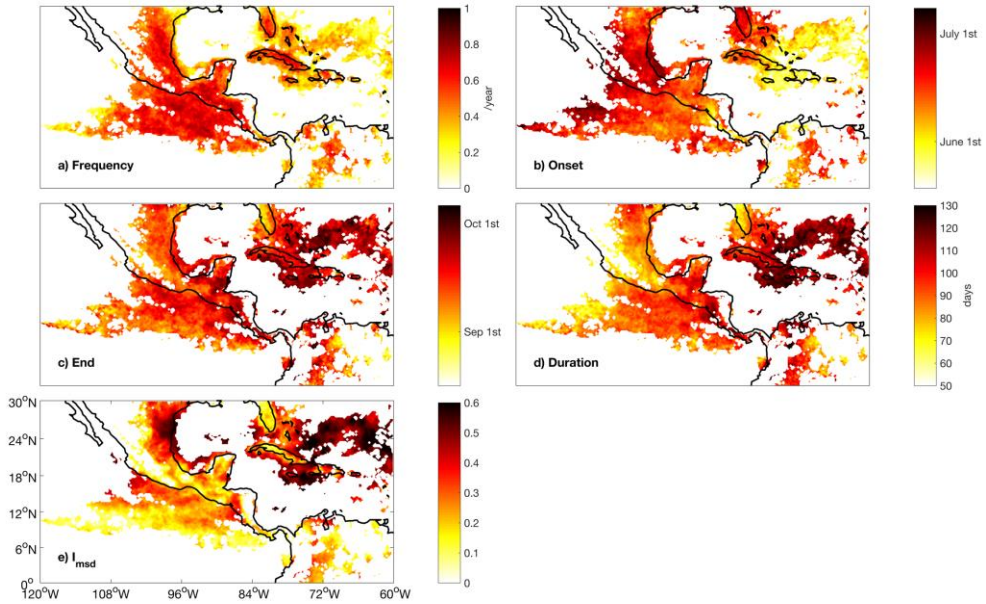

Figure 3. Metrics of MSD signatures over Central America and Mexico from TRMM. Five panels separately indicate (a) annual frequency, (b) average Julian onset dates, (c) average Julian end dates, (d) average durations and (e) climatological I_{msd}. Regions exhibiting MSD signatures are shaded by colours.

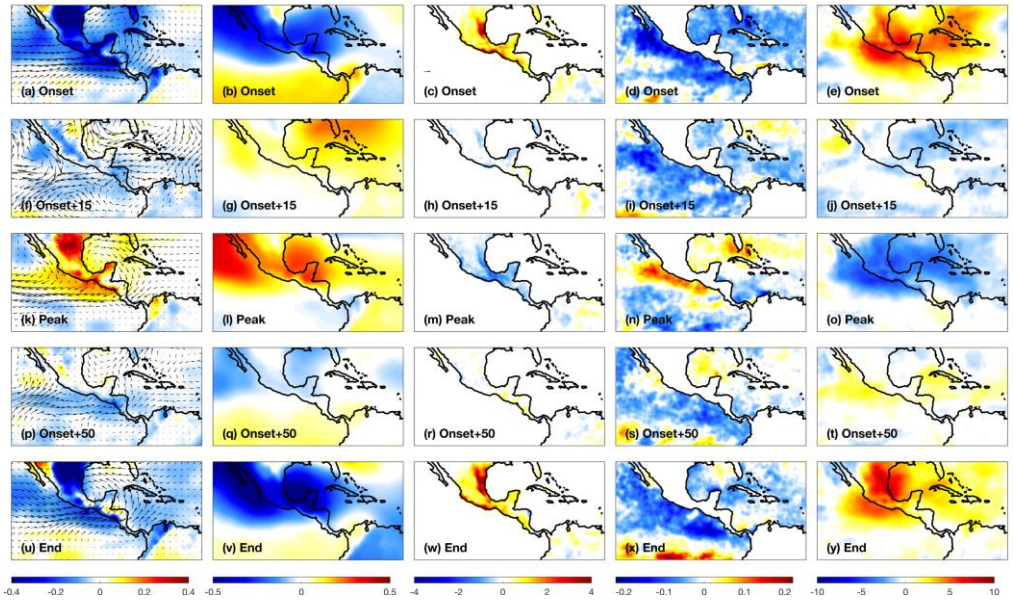

Figure 4. Dynamic propagation of atmospheric states across all detected MSDs. Each row indicates atmospheric states in a particular MSD period (onset, onset + 15 days, peak, onset + 50 days and end), while each column represents a particular atmospheric property (anomalous of 2m temperature (°C), 10m winds (m/s), surface pressure (hpa), precipitation (mm/day), sea surface temperature (°C), and total cloud fraction (%), from left to right column). Precipitation data are extracted from CPC, sea surface temperature data are extracted from NOAA OI SST V2, and others

are extracted from the ERA/L. The reference arrow of wind anomalies (1m/s) is located in Figure 4c. This reference arrow is also adapted to Figure 5, 7-10.

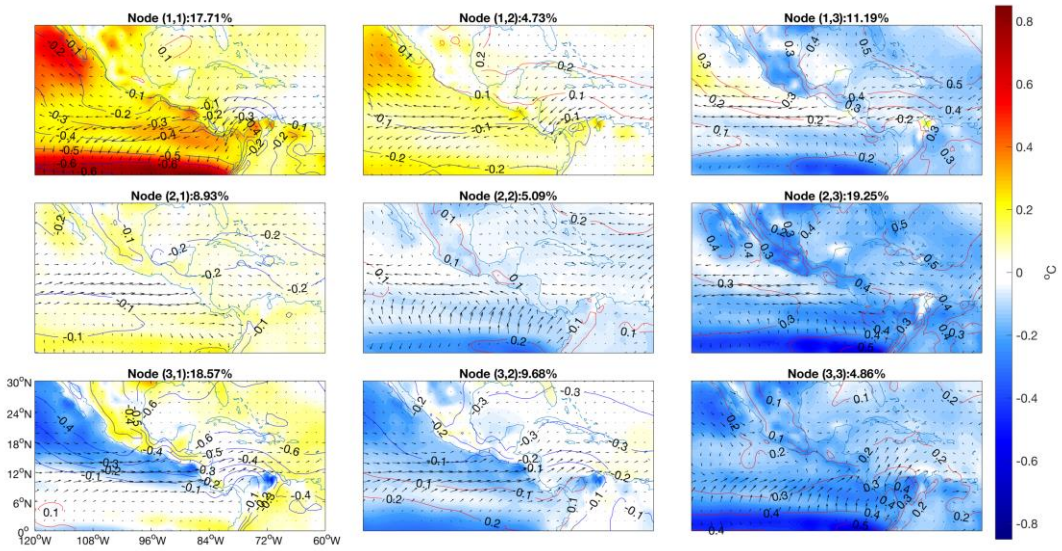

Figure 5. Atmospheric states (wind, pressure and temperature) from reconstructed SOM outputs. Colours indicate 2m air temperature anomalies and arrows indicate anomalous winds. Contours indicate the surface pressure anomalies (hPa). The percentage of MSD events assigned to each node with respect to the total count is labelled in the panel titles. Contours here are mapped for every 0.1 hPa.

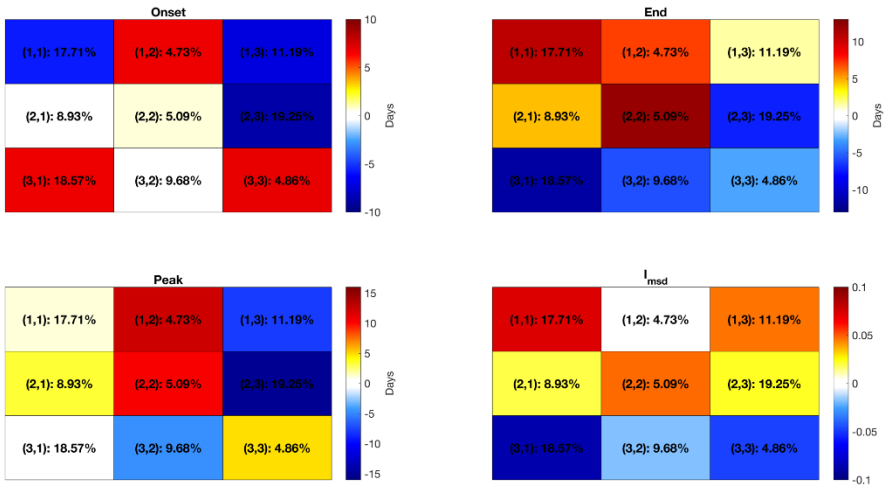

Figure 6. Averaged MSD metrics in each SOM node. Colours here indicate the difference between average MSD metrics in each SOM node and mean calculated from all MSD events. Four panels separately indicate MSD onset dates, end dates, peak dates, and I_{msd} . The percentage of MSD events assigned to each node with respect to the total count is labelled in each pixel.

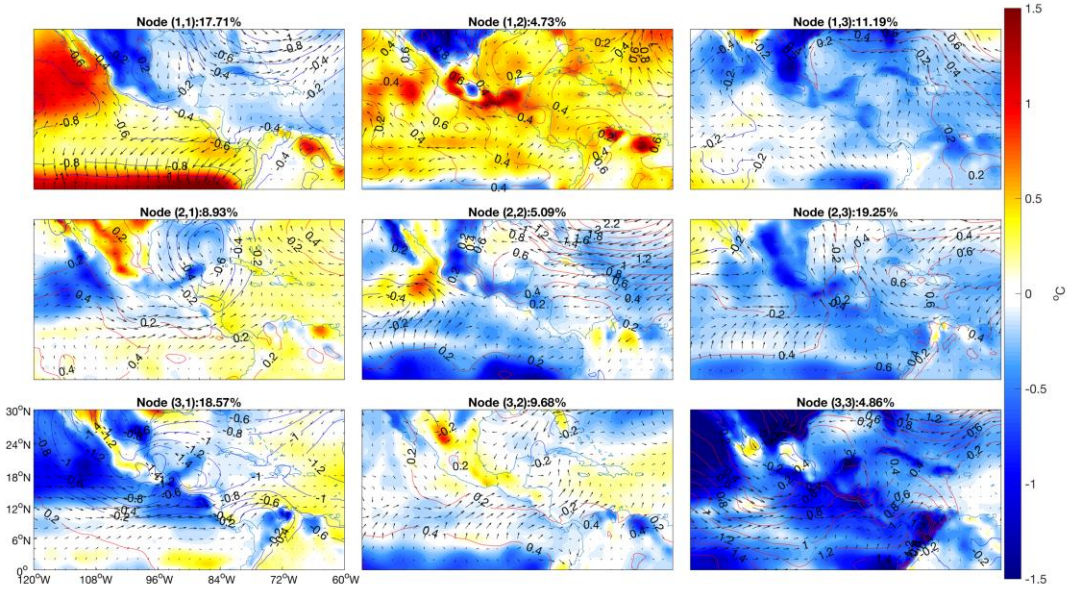

Figure 7. Average atmospheric states (wind, pressure and temperature) during onset dates of MSDs in each SOM node. Colours indicate 2m air temperature anomalies and arrows indicate anomalous winds. Contours indicate the surface pressure anomalies. The percentage of MSD events assigned to each node with respect to the total count is labelled in the panel titles. Contours here are mapped for every 0.2 hPa.

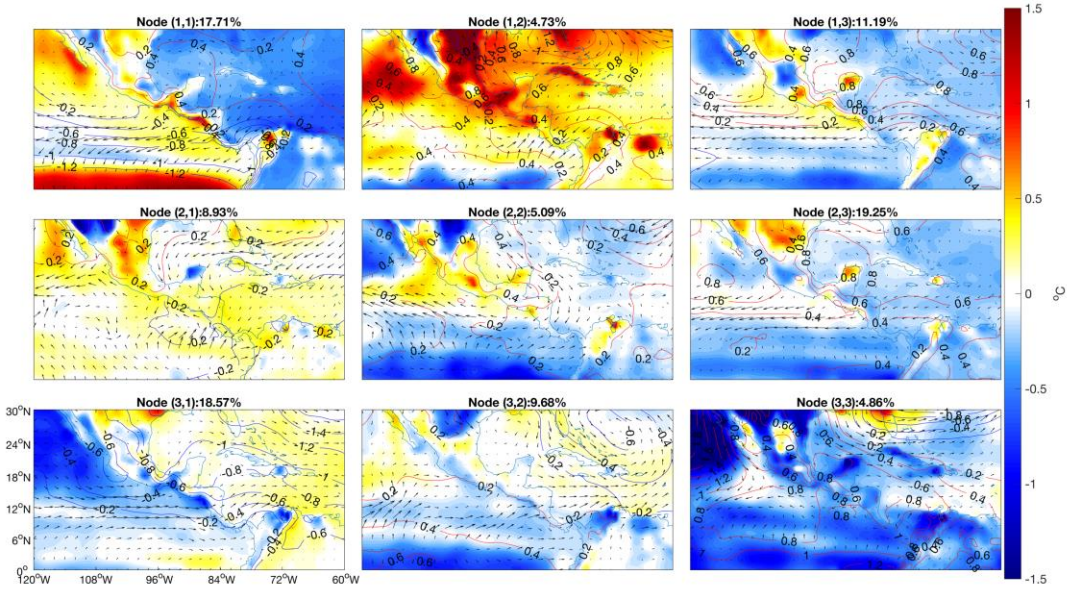

Figure 8. Average atmospheric states (wind, pressure and temperature) during peak dates of MSDs in each SOM node. Colours indicate 2m air temperature anomalies and arrows indicate anomalous winds. Contours indicate the surface pressure anomalies. The percentage of MSD events assigned to each node with respect to the

total count is labelled in the panel titles. Contours here are mapped for every 0.2 hPa.

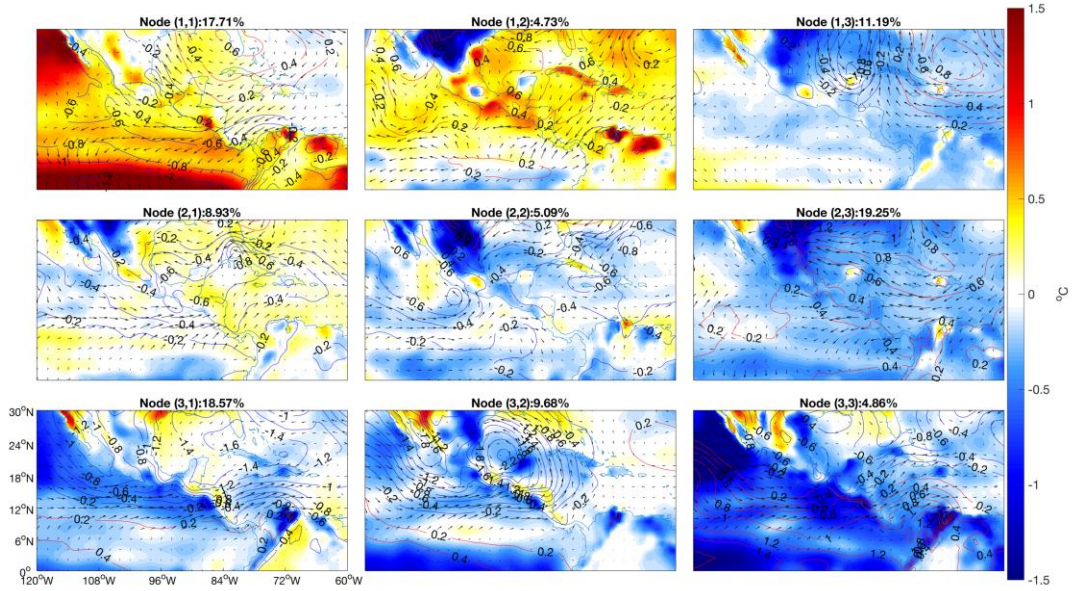

Figure 9. Average atmospheric states (wind, pressure and temperature) during end dates of MSDs in each SOM node. Colours indicate 2m air temperature anomalies and arrows indicate anomalous winds. Contours indicate the surface pressure anomalies. The percentage of MSD events assigned to each node with respect to the total count is labelled in the panel titles. Contours here are mapped for every 0.2 hPa.

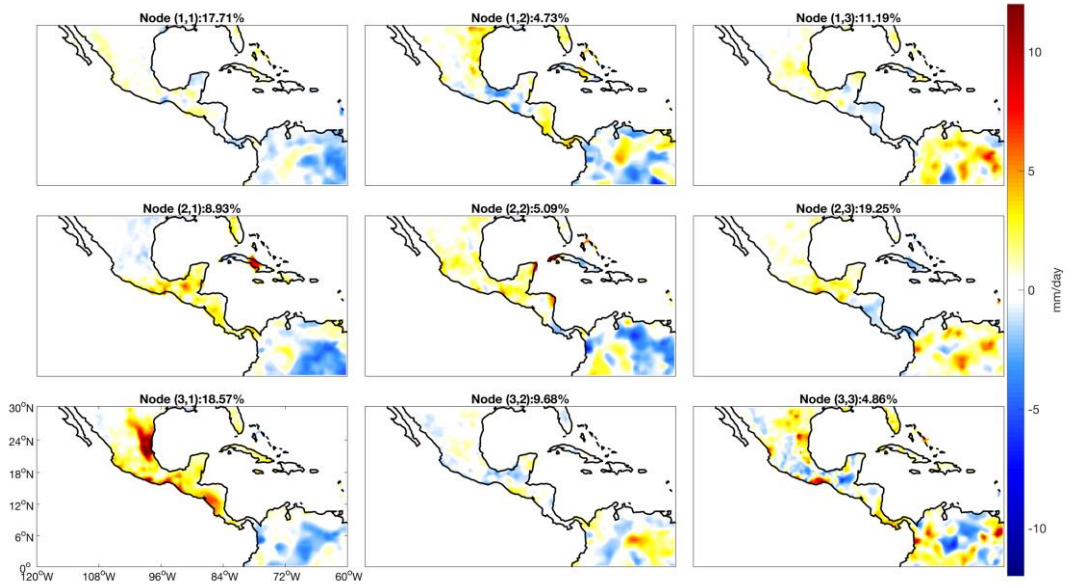

Figure 10. Average daily precipitation during onset dates of MSDs in each SOM node. Colours indicate precipitation composites in each SOM node. The percentage of MSD

events assigned to each node with respect to the total count is labelled in the panel titles.

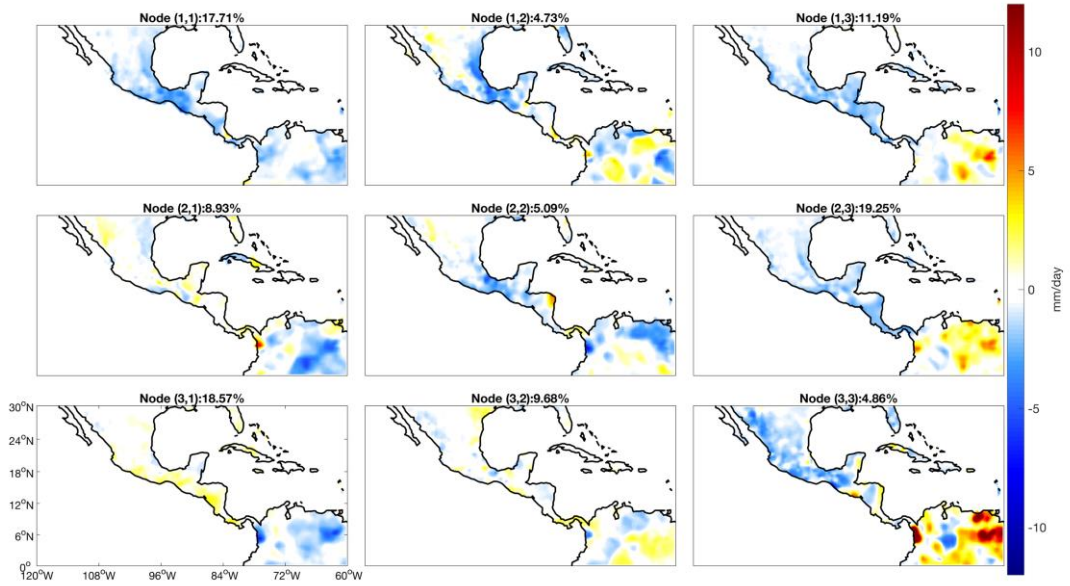

Figure 11. Average daily precipitation during peak dates of MSDs in each SOM node. Colours indicate precipitation composites in each SOM node. The percentage of MSD events assigned to each node with respect to the total count is labelled in the panel titles.

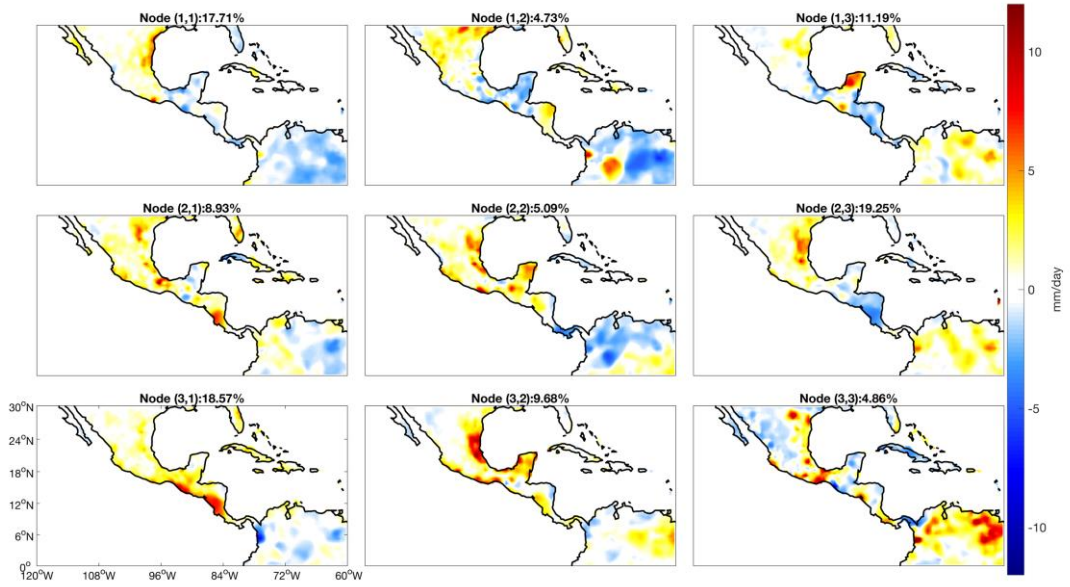

Figure 12. Average daily precipitation during end dates of MSDs in each SOM node. Colours indicate precipitation composites in each SOM node. The percentage of MSD events assigned to each node with respect to the total count is labelled in the panel

titles.

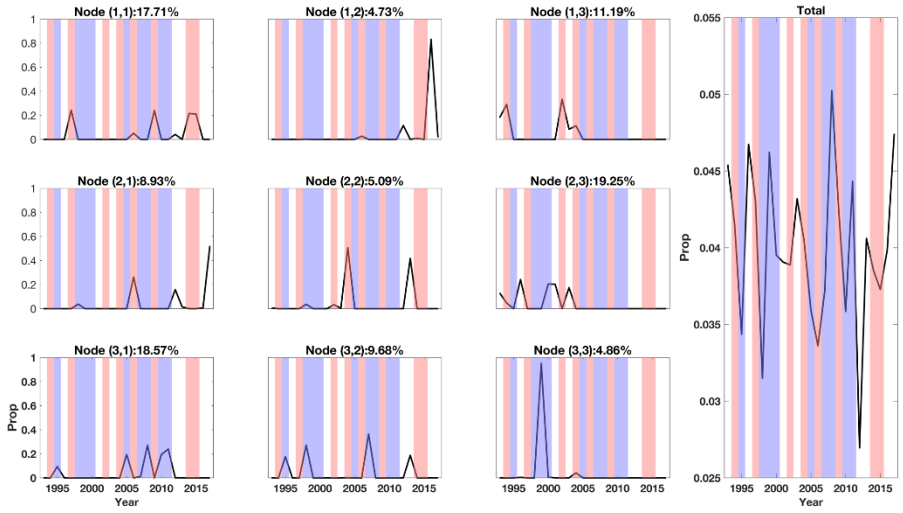

Figure 13. Occurrence time series for each SOM node and total number of MSD events. Lines indicate proportion of MSDs in each year, i.e. sum of each line should be 1. Colours indicate El Niño (red) and La Niña (blue) years.

TABLES

Table 1. Datasets used in the evaluation of proposed algorithm.

Data	Type	Grid size	Time step	Temporal Coverage	Spatial Coverage
TRMM	Observations (satellite only)	0.25° × 0.25°	Daily	1998-2017	Land and Ocean
CPC	Observations (gauge-satellite blend)	0.5° × 0.5°	Daily	1993-2017	Land
ERA/H	Reanalysis	0.25° × 0.25°	Daily	1993-2017	Land and Ocean
ERA/L	Reanalysis	0.5° × 0.5°	Daily	1993-2017	Land and Ocean

Table 2. Proportion of years and MSD signals in each ENSO mode. Row 'Year' indicates the proportion of years in each ENSO mode for 1993-2017. Row 'Total' represents the proportion of MSD signals in each ENSO mode for all detected MSD. Row '(i, j)' (i=1,2,3, j=1,2,3) indicates the proportion of MSD signals in each ENSO mode for detected MSD signals labelled in corresponding SOM node. Colours indicate corresponding proportion is statistically significant in 90% level (red; >= 90th percentile) or 10% level (blue; <=10th percentile).

	El Niño	La Niña	Neutral
Year	32.00%	36.00%	32.00%
Total	31.57%	35.51%	32.92%
(1, 1)	95.89%	0.00%	4.11%
(2, 1)	26.20%	3.73%	70.07%
(3, 1)	0.00%	99.94%	0.06%
(1, 2)	3.52%	0.12%	96.36%
(2, 2)	53.89%	3.61%	42.50%
(3, 2)	0.12%	81.16%	18.72%
(1, 3)	73.60%	0.00%	26.40%
(2, 3)	4.65%	20.41%	74.95%
(3, 3)	4.13%	95.52%	0.35%

Supplementary Material

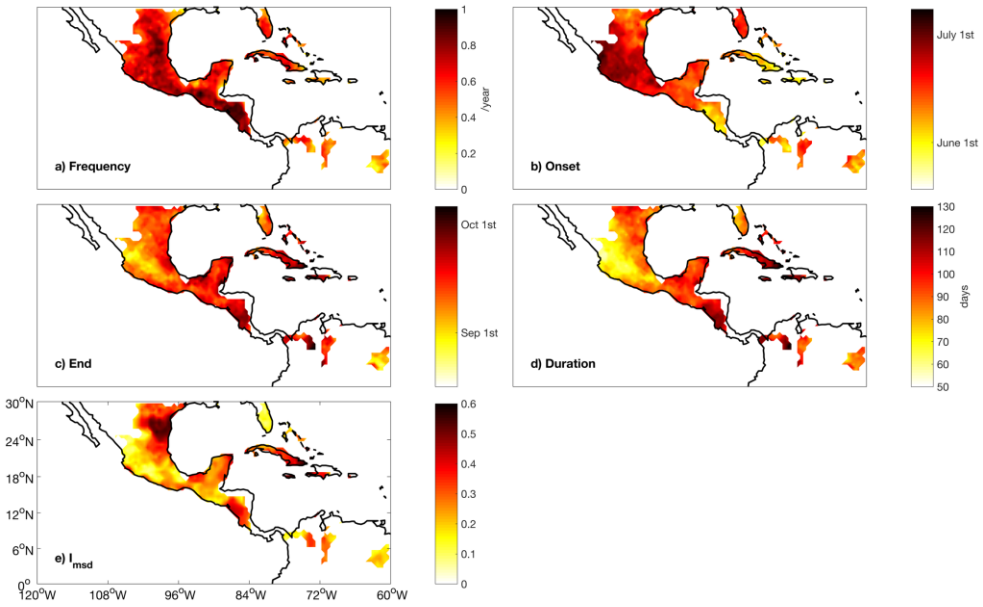

Figure S1. Similar to Figure 3, but for CPC data.

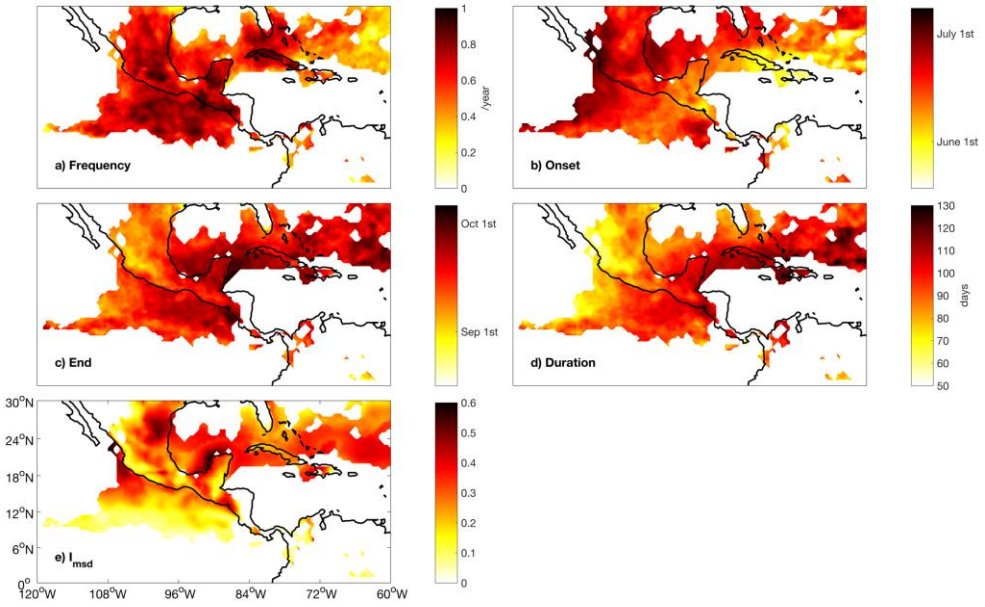

Figure S2. Similar to Figure 3, but for ERA/L data.

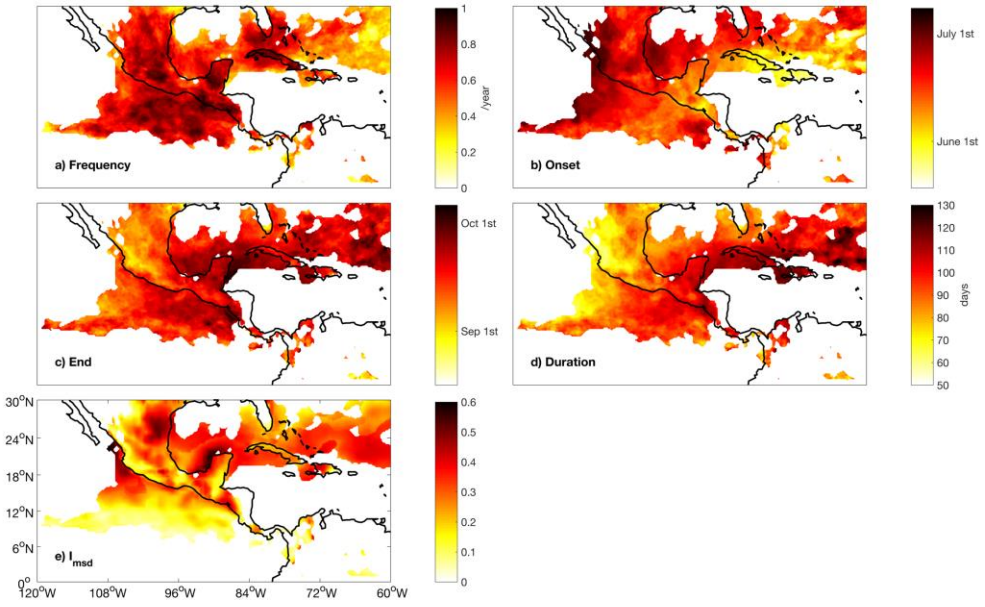

Figure S3. Similar to Figure 3, but for ERA/H data.

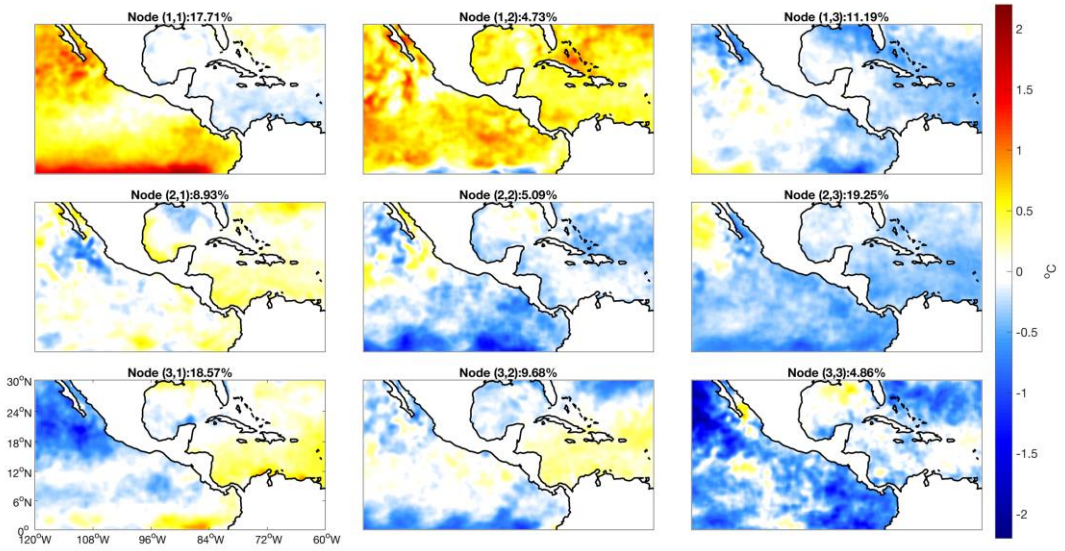

Figure S4. Average daily sea surface temperature anomalies during onset dates of MSDs in each SOM node. Colours indicate SST anomalies. The proportion of MSD events assigned to each node is labelled in the panel titles.

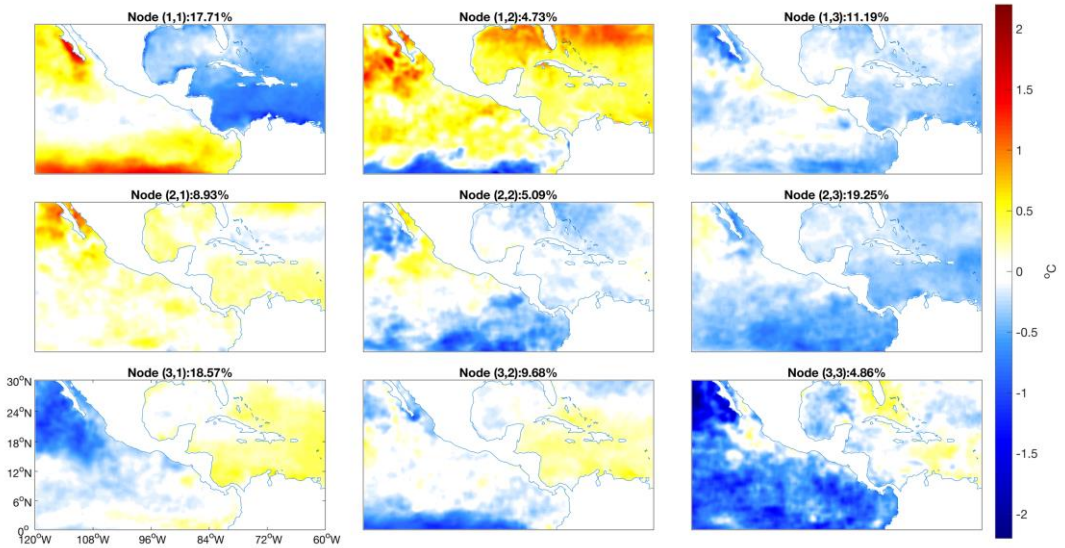

Figure S5. Average daily sea surface temperature anomalies during peak dates of MSDs in each SOM node. Colours indicate SST anomalies. The proportion of MSD events assigned to each node is labelled in the panel titles.

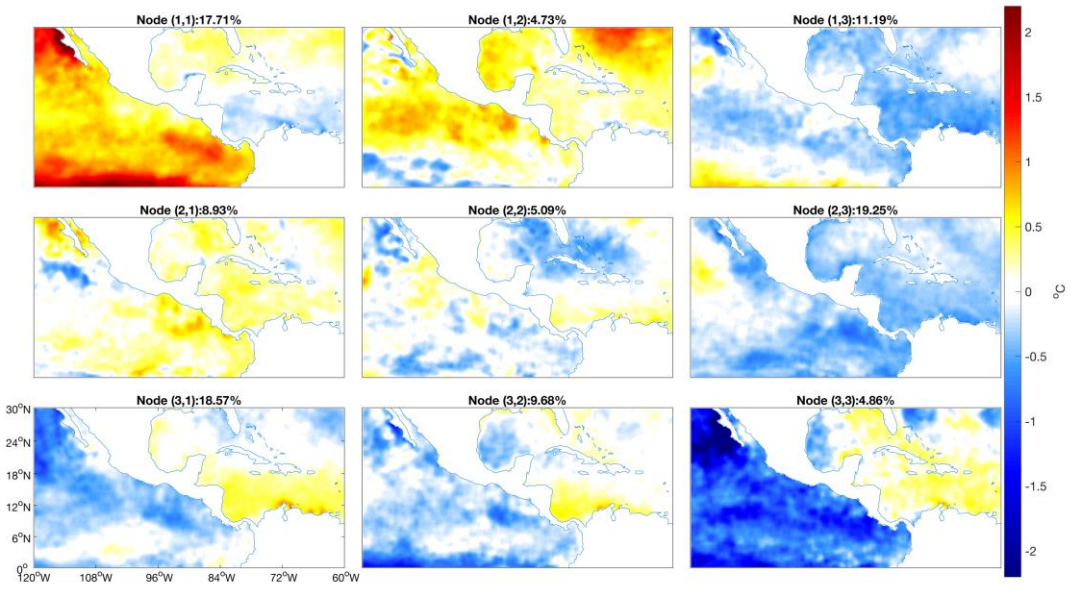

Figure S6 Average daily sea surface temperature anomalies during end dates of MSDs in each SOM node. Colours indicate SST anomalies. The proportion of MSD events assigned to each node is labelled in the panel titles.